# Ocean Color Data Merger

**Ewa J. Kwiatkowska**
**NASA Goddard Space Flight Center / SAIC**
**ewa@simbios.gsfc.nasa.gov**

The objective of ocean color data merger is to create a consistent series of systematic ocean color measurements from multi-instrument, multi-platform and multi-year observations based on accurate and uniform calibration and validation over the lifetime of the measurement.

The most obvious benefit of data merger is improvement in spatial and temporal ocean color coverage. Single sensor daily coverage is severely limited by gaps between consecutive swaths and gaps caused by clouds, sun glint and other phenomena which hinder the extraction of ocean color (Gregg *et al.*, 1998; Gregg and Woodward, 1998). For example, merged data from three global satellite sensors, MODIS on the Terra and Aqua platforms and SeaWiFS, provide only about 40% of global ocean and inland-water coverage at 9km resolution within a single day. The other critical benefit is an increase in statistical confidence in extracted bio-optical parameters. Merger algorithms can utilize sensor-varying attributes, such as spectral, spatial, temporal, and ground coverage characteristics. Merger is the ultimate tool for the creation of ocean-color climate data records.

There are many difficulties associated with ocean color data merger. Sensors have varying designs and characteristics. There are disparate instrument calibrations, data processing algorithms, and validation accuracies. The same ocean color quantities can be derived using different spectral bands and different algorithms which may cause dissimilarities in mission standard products. Discrepancies in sensor characteristics, calibrations, and data processing create relationships between data products from different instruments which may show temporal trends and dependencies on sensor observation conditions. These relationships may also be noisy, indefinite and sometimes contradictory. Data especially susceptible to noisiness are those contaminated by clouds, dust, other types of turbid atmosphere, coastal waters, and mixed pixel representations. Another type of ambiguity arises from the fact that sensors are flown over the same regions at different times of a day. Natural changes in bio-optical conditions of the global ocean occurring over these time spans are hard to establish because they are difficult to discriminate from instrument and calibration artifacts.

Detailed objectives for the creation of a consistent series of multi-instrument and multi-year ocean color observations and related ocean Climate Data Records have not yet been defined. An objective way to assess accuracy of ocean color data is through comparisons, called matchups, with *in situ* measurements (Bailey *et al.*, 2001). However, ocean color matchups against *in situ* ship-born measurements are relatively sparse. This is because of the difficulties in acquisition of *in situ* observations and uncertainties involved in comparing *in situ* measurements against satellite-derived data. Over the sensors' lifetime, there have been 250 chlorophyll-a concentration matchup points, strictly screened for

quality, for SeaWiFS and 34 for MODIS-Terra (http://seabass.gsfc.nasa.gov/matchup_results.html). Therefore, matchups with *in situ* observations are mostly used for intermittent validation of sensor data in concert with spatial and temporal data consistency analyses. The other approach to validation is matchup of ocean color data between sensors (Chapter 2; Kilpatrick *et al.*, 2001). This method assesses discrepancies between sensors over global to local zones and daily to seasonal time scales. Such assessments are vital for the data merger because they enable extraction of disparate trends and trend dependencies in data from different instruments.

There have been a number of methods developed to merge ocean color data. These methods include averaging and weighted averaging of data within the sensor overlapping coverage. Blending algorithms have been applied which fit a function over shape-of-the-field defining data from one sensor given an internal boundary condition delimited by data from the other sensor or *in situ* observations (Gregg and Conkright, 2001). A semi-analytical optical algorithm has been developed which uses combined nLw retrievals within overlapping coverage at different sensor-specific wavelengths to calculate chlorophyll concentrations, combined detrital particulate and dissolved absorption coefficients, and particulate backscattering coefficients (Maritorena *et al.*, 2002; Maritorena *et al.*, 2000).

The major data merger effort undertaken by the SIMBIOS Project Office focused on integrating ocean color data from global sensors at a daily temporal resolution. MODIS-Terra and SeaWiFS data were used to study methodologies to create a consistent series of long-term observations from sensors of different design, characterization, processing algorithms, and calibrations. The information derived from MODIS-Terra and SeaWiFS comparisons, described in Chapter 2, was used to derive an ocean color sensor cross-calibration strategy to eliminate pronounced data temporal discrepancies between the sensors and MODIS data artifacts. Statistical objective analysis was investigated to spatially and temporally interpolate MODIS-Terra (cross-calibrated with SeaWiFS) and SeaWiFS data onto daily global ocean color maps using individual sensor accuracies and producing error bars for each data point on the map. Additional research was performed to support local-area data merger applications, for instance in coastal zones. These applications utilized ocean color data of different spatial resolutions and *in situ* measurements. The multiresolution merger focused on enhancement of oceanic features in lower resolution imagery using higher resolution data.

## 1. Machine Learning Cross-Calibration of Multi-Sensor and Multi-Year Datasets To Create a Consistent and Calibrated Global Ocean Color Baseline

### *1.1 Introduction*

Described in Chapter 2 results from collection-4 MODIS-Terra and reprocessing-4 SeaWiFS data comparisons showed that over three years of concurrent operation, there were significant discrepancies in ocean color measurements between the two sensors.

Differences in water-leaving radiances and chlorophyll-*a* concentrations between the two instruments exceeded the maximum uncertainties preliminarily established for the creation of consistent merged ocean color products. These discrepancies furthermore varied temporally and spatially and were dependent on MODIS-Terra characterization problems for features such as the side of the optical mirror, response versus scan angle, and polarization sensitivity (Esaias *et al.*, 1998).

To create a consistent series of ocean color measurements from multi-instrument, multi-platform and multi-year observations, spatial, temporal and instrument artifact-driven discrepancies between sensor data had to be eliminated. To accomplish it a sensor cross-calibration approach was proposed. The goal of the ocean color sensor cross-calibration was to bring multi-instrument and multi-year data to a single well-calibrated and consistent baseline representation (Kwiatkowska, 2003). When data were cross-calibrated, daily global sensor observations could be combined to provide a joint ocean color coverage which was consistent through time and space and supported a number of applications, including the creation of ocean Climate Data Records.

Cross-calibration of ocean color sensors has been a typical approach to adjusting retrievals from instruments which do not have sufficient on-board calibration capabilities, such as Ocean Scanning Multispectral Imager (OSMI) (Franz and Kim, 2001). Sensors with operational calibrations have been also validated and cross-calibrated. Modular Optoelectronic Scanner (MOS) was vicariously calibrated using three overlapping scenes with SeaWiFS (Wang and Franz, 2000). SeaWiFS-obtained aerosol models and normalized water-leaving reflectances and MOS-derived aerosol concentration and molecular scattering estimates were used to calculate MOS top-of-the-atmosphere (TOA) radiances. For spatial fields of relatively uniform SeaWiFS TOA radiances and normalized water-leaving reflectances, MOS band calibration gains were obtained as a ratio of SeaWiFS to MOS TOA radiances. Ocean Color and Temperature Scanner (OCTS) and Polarization and Directionality of the Earth's Reflectances I (POLDER) sensors were vicariously recalibrated with common *in situ* nLw data and a consistent atmospheric correction to produce relatively comparable ocean color products with no obvious bias differences (Wang *et al.*, 2002). With more complex ocean color sensors currently on orbit, such as MODIS, the cross-calibration task is more difficult and a higher temporal and spatial accuracy of ocean retrievals is expected.

In this implementation, SeaWiFS was consequently chosen as the ocean color baseline data set because it had a long history of calibration and validation efforts (Barnes *et al.*, 2001; Eplee *et al.*, 2001), its data proved stable and self-consistent through the years, and it was invariably used as a standard against which other sensors were cross-calibrated. The cross-calibration aimed to bring MODIS-Terra product data to SeaWiFS-like values or, in other words, to emulate SeaWiFS response given MODIS data. This was accomplished by deriving a comprehensive machine learning approach. Machine learning methodology was attractive because the amount of required a priori knowledge about detailed sensor characterization, response changes, and the uncertainties of the radiative transfer modeling, calibration, and algorithms, was minimal. The machines learned from examples using abundant data from global overlapping coverage between MODIS and

SeaWiFS. They worked as regularity detectors to discover statistically salient properties of investigated data.

Machine learning techniques have been used extensively in ocean color data processing. Various sensor calibration and atmospheric correction parameters were derived by means of regression (Barnes *et al.*, 2001; Eplee *et al.*, 2001; Gordon and Wang, 1994). The empirical algorithm associating chlorophyll concentration with water-leaving radiances was a result of the polynomial regression using *in situ* measurements (O'Reilly *et al.*, 1998). Multivariate optimization techniques were applied to ocean color data to simultaneously retrieve in-water biophysical conditions and aerosol optical properties in atmospheres containing weakly and strongly absorbing aerosols (Gordon *et al.*, 1997; Chomko and Gordon, 1998).

The cross-calibration strategy employed here was based on two foundations. Firstly, it defined a variety of MODIS data products and parameters necessary to alleviate the effects of MODIS temporal and spatial trends and artifact-driven dependencies in data. Secondly, large amounts of joint MODIS and SeaWiFS data were applied to serve as examples of these dependencies in order to determine the most appropriate regression functions from MODIS to SeaWiFS data. Three types of machine learning techniques were used in the current implementation: evolutionary computation, clustering, and regression.

*1.2 Machine Learning Techniques for Cross-Calibration*

Machine learning techniques were employed to define MODIS inputs to the cross-calibration which were essential in resolving the temporal and spatial trend and artifact-driven dependencies in MODIS data. The space of possible MODIS data products and parameters was searched for the most effective set of features enabling data dependence decorrelation and cross-calibration with SeaWiFS. These features needed to provide information on MODIS sensor and measurements that most unambiguously defined MODIS data and made possible one-to-one mapping to SeaWiFS. A set of possible features encompassed a variety of information describing MODIS data including water-leaving radiances at all visible ocean bands, chlorophyll-*a* concentration, K_490, atmospheric parameters such as AOT and ε, MODIS viewing and solar geometry, geographical situation, day in the time sequence, and ancillary meteorological setting of each MODIS data point. Conventional statistical and non-parametric rank-statistical correlation algorithms (Press *et al.*, 1992) were found insufficient to extract dependencies in MODIS data because of the nonlinearity and fuzziness of the cross-calibration problem. Consequently, an evolutionary-computation search mechanism was applied through a genetic algorithm (Goldberg, 1989). The genetic algorithm evaluated and propagated the fitness of various combinations of MODIS feature inputs through generations of non-parametric regression neural networks which mapped these inputs to SeaWiFS chlorophyll. The neural networks were trained on a multi-day MODIS and SeaWiFS data set which was scaled down for fast processing.

Clustering was employed in the cross-calibration process to partition the feature space of multi-day, multi-year, global overlapping MODIS-Terra and SeaWiFS data into clusters of similar feature values. Separate cross-calibration processes were then performed on each cluster data. This limited the complexity and improved the accuracy of the overall sensor cross-calibration. In this implementation, a Linde-Buzo-Gray clustering algorithm was applied (Linde *et al.*, 1980).

Regression was the actual tool that determined the relationship between MODIS and SeaWiFS data needed by the cross-calibration. The pre-merger validations presented in Chapter 2 showed that the cross-calibration problem was highly nonlinear and highly multidimensional with many dependencies present among data variables. Therefore, the cross-calibration depended closely on a suitable choice of MODIS input features, adequate clustering which spanned the entire space of MODIS and SeaWiFS feature data, and on the regression algorithm, which needed to be effective with highly nonlinear and ambiguous problems. Although the mapping between MODIS and SeaWiFS data could be performed using conventional linear or non-linear regression, the use of artificial neural networks or support vector machines was preferred (Pao, 1989). Neural networks and support vector machines could deal with flawed, biased, and cross-dependent sensor data because they are distribution free and can support highly nonlinear decision boundaries in the feature spaces.

Artificial neural networks have been widely employed in remote sensing, mainly however to solve classification problems, such as land cover or cloud type categorization (Atkinson and Tatnall, 1997; Ainsworth and Jones, 1999; Gross *et al.*, 2000; Tanaka *et al.*, 2000). In the current implementation, neural networks were initially employed to perform the regression between MODIS and SeaWiFS data (Kwiatkowska and Fargion, 2002a). However, with the training sets becoming massive and encompassing a large number of MODIS input features and multi-day, multi-year, global joint sensor coverages, the neural networks turned out too slow in training. Support vector machines replaced the neural network regression. Support vector machines are learning kernel-based systems that use a hypothesis space of linear functions in high dimensional feature spaces (Cristianini and Shawe-Taylor, 2000). Unlike neural networks, which try to define complex functions in the input feature space, the kernel methods perform a non-linear mapping of the complex data to high dimensional feature spaces and then use linear functions to create decision hyperplanes or regression functions (Schölkopf, 2000). The problem of choosing an architecture for a neural network is replaced by the problem of choosing a suitable kernel. Kernel functions project the data into high dimensional feature spaces to increase the computational power of the linear learning machines. The advantages of support vector machines over neural networks are that they train significantly faster, are better suited for work with high dimensional data, they often have a single minimum to search, and they allow for scaling the importance of outliers. Support vector machines have been applied in remote sensing to solve classification problems (Azimi-Sadjadi and Zekavat, 2000; Gualtieri *et al.*, 1999; Fukuda and Hirosawa, 2001). Applications of support vector machines in remote sensing to solve regression problems are sparse (Brown *et al.*, 2000) and there are no reported implementations involving ocean color data.

*1.3 Machine Learning Implementation*

The ocean color merger approach defined here was aimed at minimizing spatial and temporal discrepancies between MODIS and SeaWiFS data, and scan angle and latitudinal dependencies in MODIS data to produce consistent daily global coverages from both sensors. The goal of the sensor cross-calibration was to reproduce uniform SeaWiFS baseline response from MODIS data. To obtain the cross-calibration, support vector machines were used to approximate the regression functions from examples of MODIS and SeaWiFS concurrent coverages. Support vector machines learned complex relationships between MODIS and SeaWiFS ocean data and were required to extend this knowledge to unknown cases.

In this implementation, the regression was obtained between MODIS data and SeaWiFS chlorophyll to merge both sensor data into a time series of daily global chlorophyll baseline data sets. However, the regression could also be performed to generate baseline data sets for water-leaving radiances and other products. The regression functions were extracted using data from daily overlapping global bin coverages between MODIS-Terra and SeaWiFS over a significant time series. L3 binned data were used at 36km resolution. The 36km bin size was sufficient to extract generalized temporal and spatial trends and instrument and calibration artifacts from MODIS data. The total data set encompassed 44 days of joint MODIS and SeaWiFS coverage spanned through time from October 2000 to July 2002. This set was comprised of 1,672,188 examples and was subsetted in various ways into training and testing sets for the support vector machine learning. The results presented in this chapter were obtained by training on 42 days of overlapping MODIS and SeaWiFS coverage, where, for each day, data from alternating bins were placed into the training and testing sets. The remaining two days where used as an additional testing set to evaluate the machine's performance at extrapolating its knowledge through time to unknown days of data.

Before applying the regression, input data had to be prepared to facilitate the support vector machine training. Elimination of linear trends in data, seasonal components, and slow variation could be important data preprocessing tasks because the algorithm might ignore decisive subtle information present in the data in favor of large variations exhibited by a trend (Masters, 1994). The seasonal trends, North-to-South variations, and MODIS scan-angle dependencies shown by MODIS and SeaWiFS comparisons in Chapter 2 were however ambiguous and intertwined. For the current implementation, none of these deterministic components were removed from data. Instead, the regression was made to exploit indirect information about these conditions included in input feature data. The other data preprocessing consideration was scaling. Global chlorophyll concentration inherently forms lognormal probability density functions (Campbell, 1995). MODIS and SeaWiFS chlorophyll feature data were therefore passed through the logarithm function to provide the most effective spread of their distributions. All input feature data, including the chlorophyll logarithms, were then scaled and translated so that they were within the limits from 0 to 1 in value. The scaling simplified the regression

because data values were then distributed within a small known range and encouraged equitable spread of importance among input feature elements.

The genetic algorithm found 16 MODIS input features to be the most effective in decorrelating MODIS data dependencies and these features were consequently used to perform the cross-calibration from MODIS to SeaWiFS data:

- nLw_412, nLw_443, nLw_488, nLw_531, nLw_551, nLw_678,
- chlor_a_2,
- Tau_865,
- Eps_78,
- satellite zenith angle,
- solar zenith angle,
- longitude and latitude,
- ozone amount,
- humidity,
- and date.

After initial regression training exercises, it was determined that a single support vector machine would be slow in learning decision functions on large data sets. Extensive data sets were however needed to present the regression algorithm with a large variety of possible feature distributions and relationships between MODIS and SeaWiFS data, especially considering the noisiness and fuzziness of these relationships and complex dependencies in data. To increase the efficiency and the accuracy of the regression, input training data were clustered and then separate support vector machines were trained on their corresponding data clusters. From plots and statistics it was established that the clusters adequately spread data distributions for all input features. The distributions were dependent on feature data probability densities within the training set. Parallel training on smaller data sets decreased support vector machine learning times by a few orders of magnitude, depending on the numbers of clusters and sizes of the training sets. Applying multiple support vector machines trained on their individual clusters of data increased the accuracy of the regression, which was critical to make the cross-calibration useful. One thousand twenty four (1024) clusters were used to obtain the results presented in this chapter, although different numbers of clusters were also investigated.

The regression emulated the response from the SeaWiFS baseline given data from MODIS. To obtain the regression-based cross-calibration, support vector machines were trained on clusters of MODIS feature data to reproduce corresponding SeaWiFS chlor_a values. There was one support vector machine for each cluster. The machines discovered the relationships between MODIS data and SeaWiFS chlorophyll and stored them in their support vectors and coefficients. The training was automatic and once accomplished, the stored vectors and system parameters could produce regression output for new data with no significant processing. The convergence of support vector machines depended on the choice of a kernel function. This study found the radial basis function to perform better than linear, polynomial, and ANOVA kernels in support of regression between MODIS and SeaWiFS data (Schölkopf *et al.*, 1996). The results presented later in this chapter were obtained using radial basis kernels:

radial basis function = $e^{-\gamma\|x-y\|^2}$,
where the γ parameter was equal to 1.0. A capacity parameter used in support vector machines measured the richness or flexibility of the regression functions and provided the protection against overfitting. An initial search for the capacity parameter, which improved the generalization accuracy of the learning system, resulted in selecting the capacity value of 8.4 for the subsequent investigations. Another parameter used in the training of support vector machines was ε. ε controlled the insensitivity of the regression by setting the predictions which lied within ε distance of their true values to be sufficiently accurate (Hearst, 1998). ε was chosen to be equal to 0.001.

*1.4 Machine Learning Cross-Calibration Results*

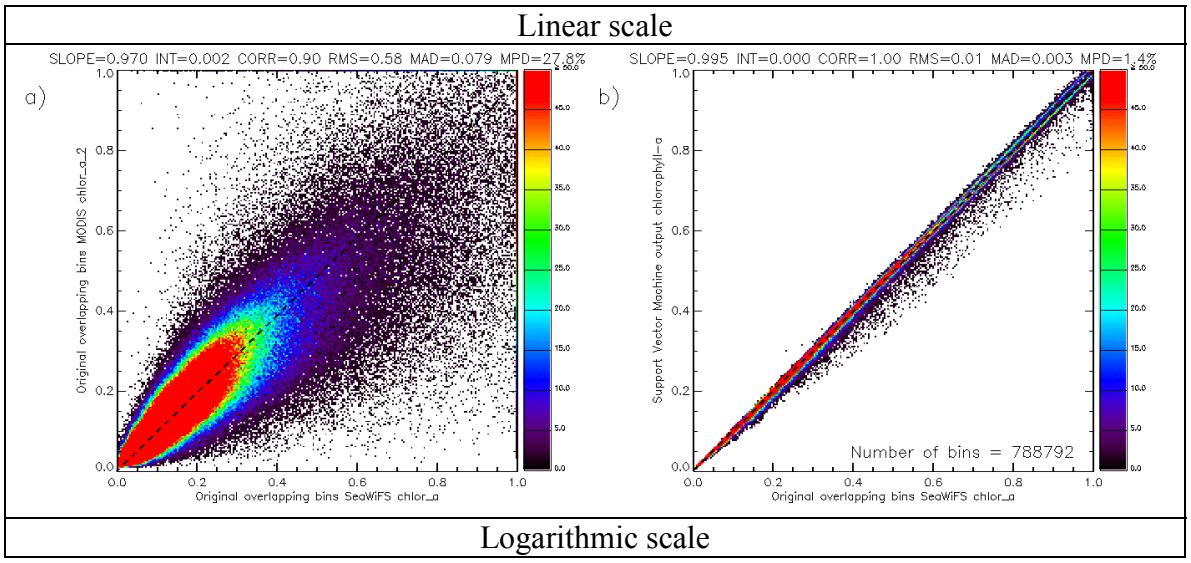
The MODIS and SeaWiFS cross-calibration approach evolved over dozens of regression training trials initially using neural networks and then support vector machines, applying different parameters and data sets for training and testing. Some of these trials were described in Kwiatkowska (2003). The results presented here were obtained by a recent version of the cross-calibration strategy.

A large set comprising 42 days of global overlapping MODIS and SeaWiFS data was used to create a representative training set for the support vector machine sensor cross-calibration. The dates spanned in time from October 2000 to July 2002 in about half-a-month intervals. For each day, data from every second overlapping bin between the sensors were placed into the training set and the remaining data were included in the testing set. There were 788,792 examples in the training set and 788,772 examples in the testing set. The additional testing set comprised complete 2 days of global overlapping data, which were separate from the 42 days used in training, and contained 94,624 examples. The training data was clustered into 1024 clusters and individual support vector machines were trained on each cluster.

To evaluate the regression, scatter plots, called matchups, were created for SeaWiFS chlorophyll versus the result of support vector machine regression from MODIS data. These plots were presented alongside the scatter plots of SeaWiFS versus original MODIS chlorophyll. Corresponding statistics were calculated for both data matchups and displayed with the plots. The statistics were used to quantify the improvements introduced by the machine learning cross-calibration to the time series of MODIS chlorophyll distribution in comparison with the SeaWiFS baseline. The statistics included slope (SLOPE) and intercept (INT) of the linear fit between SeaWiFS and MODIS data. To calculate the linear fit, an outlier-resistant linear regression was applied based on the robust Tukey's biweight calculated perpendicularly to the bisector of MODIS vs. SeaWiFS and SeaWiFS vs. MODIS data (Press *et al.*, 1992). The robust bisquare weighting ensured that the slope and intercept parameters were representative of the bulk of the data distribution and were not skewed by a few outlier points. Furthermore, matchup data robust correlation (CORR), root mean squared error (RMS), mean absolute

difference ( $\mathbf{MAD} = \frac{\sum_{i=0}^{n-1} \left| x_{MODIS_i} - x_{SeaWiFS_i} \right|}{n}$ ), and mean percentage difference ( $\mathbf{MPD} = 100\% \cdot \frac{\sum_{i=0}^{n-1} \frac{\left| x_{MODIS_i} - x_{SeaWiFS_i} \right|}{x_{SeaWiFS_i}}}{n}$ ) were obtained. Statistics were calculated in the linear space of chlorophyll data. The linear space allowed an effective scrutiny of the MODIS and SeaWiFS cross-calibration statistics within most prominent ocean provinces, including low chlorophyll waters and the gyres, and high chlorophyll coastal zones. Because the chlorophyll probability density had a lognormal distribution (Campbell, 1995), matchups and corresponding statistics were also displayed in the logarithmic chlorophyll space, except for the MPD which was calculated in the linear space.

Support vector machines were accurately trained on the 42-day training set. Fig. 1 shows a scatter plot of original SeaWiFS and MODIS chlorophyll values from overlapping daily coverage contained in the training set, a), and a scatter plot of original training SeaWiFS chlorophyll against the chlorophyll obtained by regression from MODIS data, b). In an ideal case, data within plot b) should lie on the y=x line and the training data are very close to this distribution. The support vector machines therefore learned the training set almost perfectly.



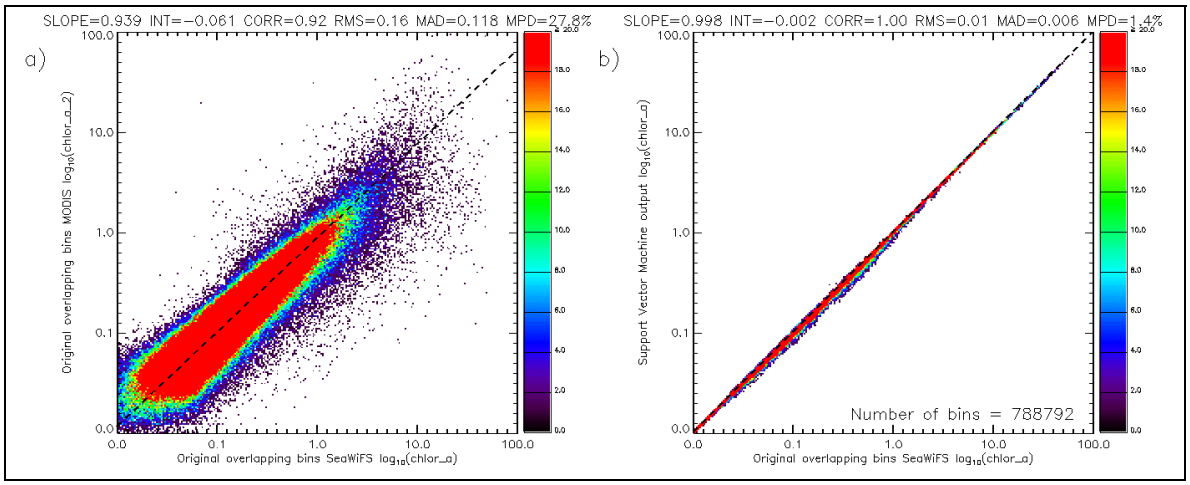



Fig. 1. Scatter plots for the 42-day training data set a) SeaWiFS against MODIS training set chlorophyll and b) SeaWiFS training set chlorophyll against chlorophyll regressed from MODIS data. The support vector machines were trained highly accurately to reproduce SeaWiFS chlorophyll data.

The support vector machines were then evaluated against the 42-day testing set data. The result of the evaluation is displayed in Fig. 2. Fig. 2b shows that the cross-calibration introduced substantial improvements to the general trends in MODIS chlorophyll distribution when compared to the original MODIS and SeaWiFS discrepancies from Fig. 2a. The improvements are present in all statistical parameters obtained from the matchups, including enhanced slope and intercept of the linear fit in data, increased correlation, and decreased mean absolute differences. The mean percent difference between MODIS and SeaWiFS chlorophyll was decreased by over 10% across the two-year daily global time series. For the purpose of a consistent series of multi-instrument and multi-year ocean color observations, this decrease in overall sensor discrepancies was significant. Although the cross-calibration demonstrated the ability to correct the distribution of the majority of MODIS chlorophyll, the support vector machines also misclassified some examples which appear in plot b) as noise away from the y=x line. The low density of the misclassified points indicated that they were infrequent. The misclassifications were caused by the overall noisiness of the data set, the complexity and inconsistency of sensor data relationships, and the difficulty to establish a fine boundary between regression function overfitting and overgeneralization.

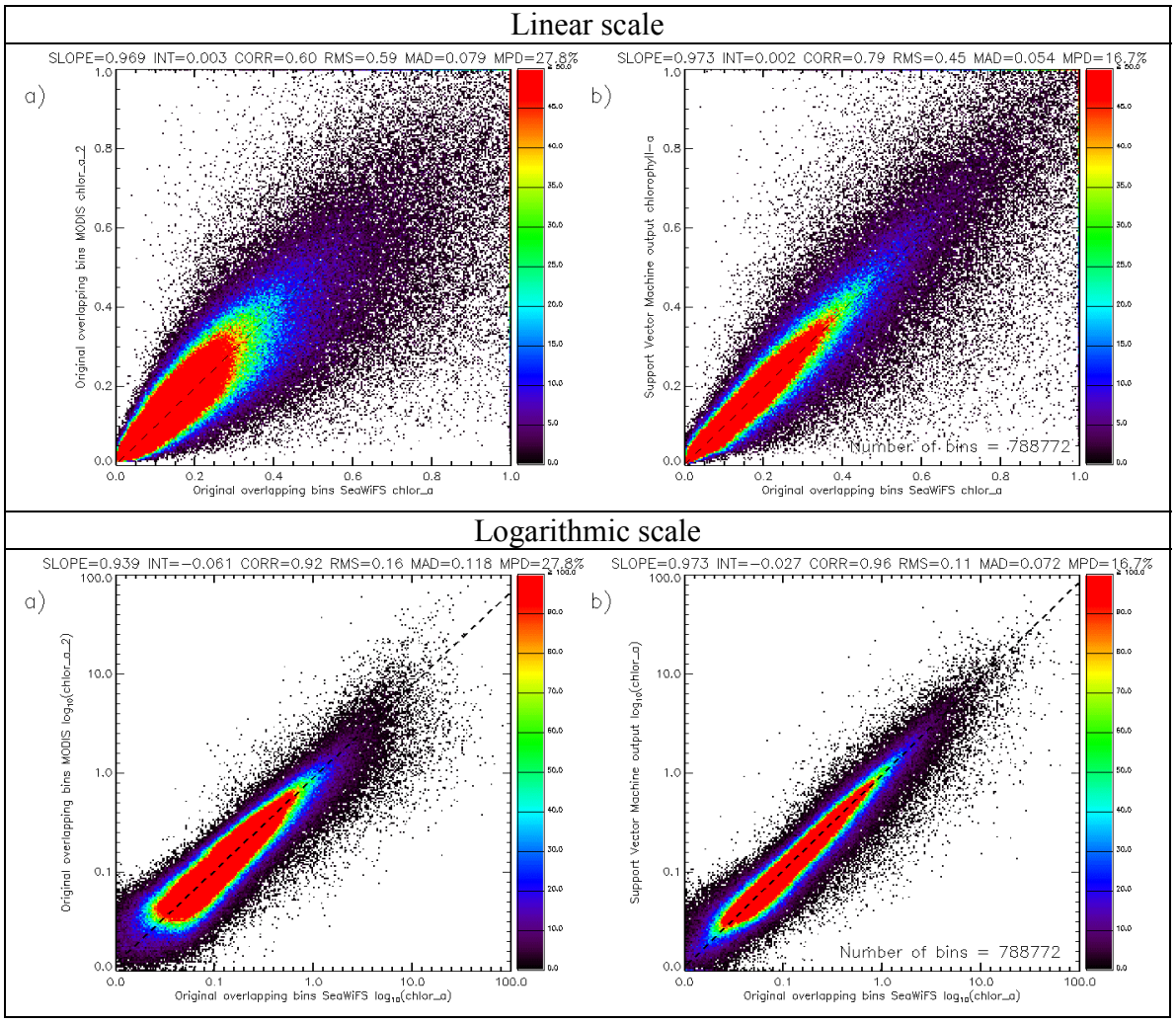



Fig. 2. Scatter plots for the 42-day testing data set: a) SeaWiFS against MODIS testing set chlorophyll, and b) SeaWiFS testing set chlorophyll against chlorophyll regressed from MODIS data. The support vector machine regression substantially improved the distribution of the bulk of MODIS chlorophyll data compared to SeaWiFS chlorophyll.

The significance of the improvements obtained through MODIS and SeaWiFS cross-calibration was demonstrated by the reduction of artifacts present in MODIS data. To evaluate the changes in MODIS data distribution in more detail, the regression's impact on MODIS data temporal trends and scan-angle and latitudinal dependencies was investigated. The testing results of support vector machine cross-calibration were separated into the individual days and analyzed independently. For each day, matchups were created between original MODIS and SeaWiFS testing chlorophyll and between SeaWiFS chlorophyll and the result of the support vector machine regression. The MPD and the slope of the linear fit gave a good estimation of the discrepancies between the corresponding data sets (Chapter 2) and were used to evaluate the performance of the mapping.

Fig. 3 illustrates MPDs, calculated against the SeaWiFS benchmark, plotted for each day and connected by lines through 42 days of the time series. The MPDs were obtained between SeaWiFS and MODIS chlorophyll and between SeaWiFS chlorophyll and the result of the regression mapping from MODIS data. Throughout the time series, the MPDs are consistently and substantially smaller for the support vector machine chlorophyll than for MODIS chlorophyll. Fig. 3 also shows that the MPDs between SeaWiFS and original MODIS chlorophyll vary seasonally. These seasonal sensor chlorophyll disparities disappear in the MPD plot for SeaWiFS and the regression result. The support vector machine cross-calibration therefore eliminated the seasonal trends in MODIS and SeaWiFS data discrepancies and decreased the differences in chlorophyll data between the sensors.

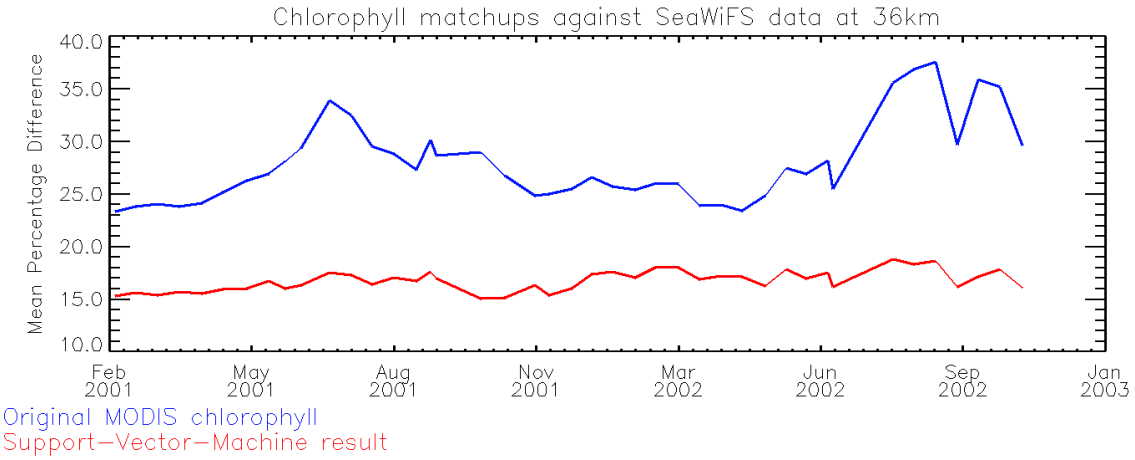



Fig. 3. Time trends in daily mean percent differences between SeaWiFS and original MODIS chlorophyll and between SeaWiFS chlorophyll and the result of the support vector machine regression from MODIS data.

From MODIS and SeaWiFS comparisons of daily global data sets described in Chapter 2 it was established that individual matchups with SeaWiFS for MODIS western and eastern scan-part data revealed distinct patterns in MODIS data behavior at different scan angles. To investigate the change in MODIS scan angle dependencies obtained by MODIS and SeaWiFS cross-calibration, MODIS daily global data were divided into subsets corresponding to their scan angle coverages and the subsets were individually matched against SeaWiFS chlorophyll. The two subsets of main concern corresponded to data located on the western and eastern scan edges of the MODIS swath. The scan edge widths constituted almost $40^0$ of MODIS zenith angle coverage for each part of the scan. For each day within the 42-day time series, MODIS chlorophyll data from the testing set, located within these two scan-edge coverages, were separately matched against SeaWiFS test set chlorophyll. Similarly, the output of the support vector machine tests corresponding to MODIS western and eastern scan edges was compared against SeaWiFS test set chlorophyll. It was demonstrated in Chapter 2 that slope and intercept of the linear fit between the matched data gave a good estimate of scan angle dependencies in MODIS products. Fig. 4 contains plots of slope values obtained in the two comparisons between testing matchup data coinciding with MODIS western and

eastern scan-edge coverages. To produce the figure, the slope values were connected across the 42-day time series into slope lines. Scan angle dependencies in original MODIS chlorophyll appear in Fig. 4a as systematically different slope values for western and eastern MODIS-scan test data in matchups with SeaWiFS. The slope lines obtained from matchups against support vector machine testing results in Fig. 4b run almost conjointly. This provided evidence that MODIS scan angle dependency was adequately eliminated from the regression result by the sensor data cross-calibration.

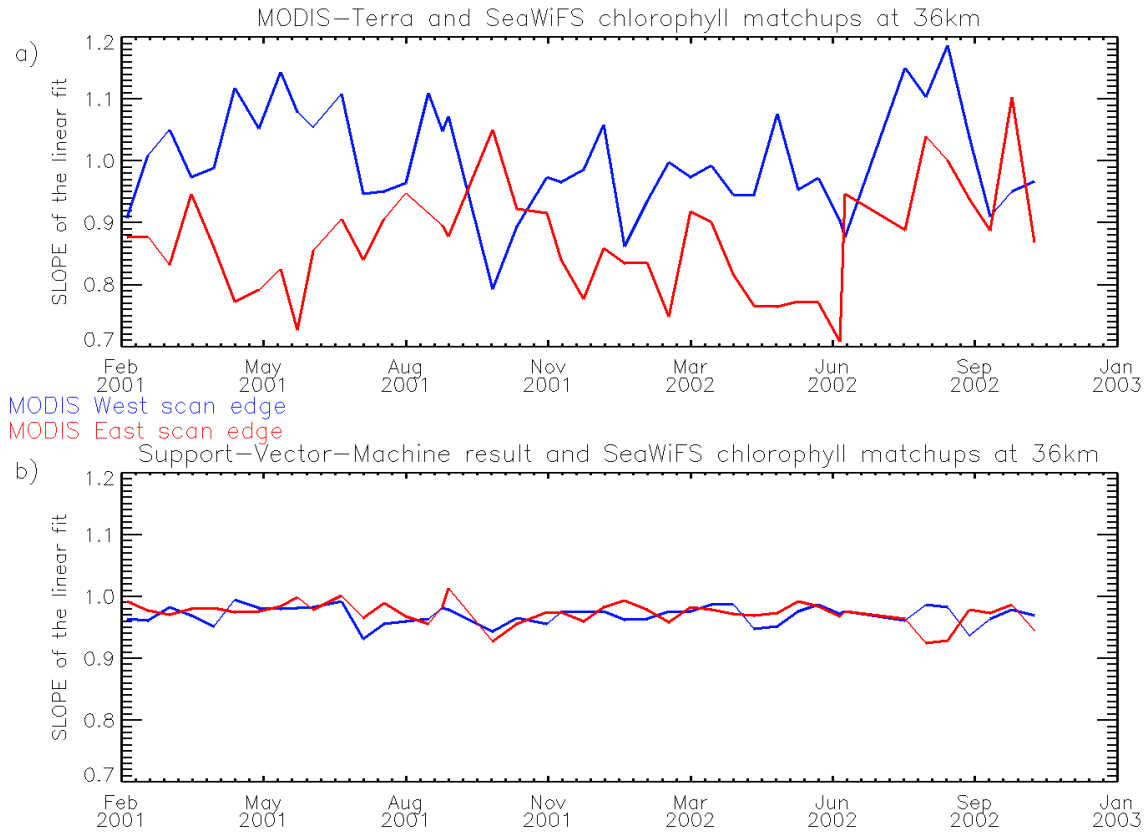



Fig. 4. Time trends in the slope of the linear fit between a) SeaWiFS and original MODIS chlorophyll for separate western and eastern MODIS scan-angle coverages and b) between SeaWiFS chlorophyll and the result of the support vector machine regression from MODIS data for the same MODIS scan coverages.

Comparisons of daily global MODIS and SeaWiFS data described in Chapter 2 demonstrated that sensor data discrepancies exhibited latitudinal dependencies. To investigate change in the latitudinal pattern produced by MODIS and SeaWiFS cross-calibration, daily global testing set data were divided between the northern and southern hemispheres and the two coverages were investigated independently. Fig. 5 illustrates slope in the linear fit between SeaWiFS and original MODIS chlorophyll and between SeaWiFS and the result of support vector machine regression obtained separately for the northern and southern hemispheres over the 42-day time series of testing set data. Fig. 5 demonstrates that, while the northern-to-southern hemisphere discrepancies were very prominent in original chlorophyll data in plot a), the discrepancies largely disappeared

from the cross-calibrated data in plot b). Support vector machine regression therefore eliminated latitudinal dependencies in MODIS data in comparison with SeaWiFS chlorophyll.

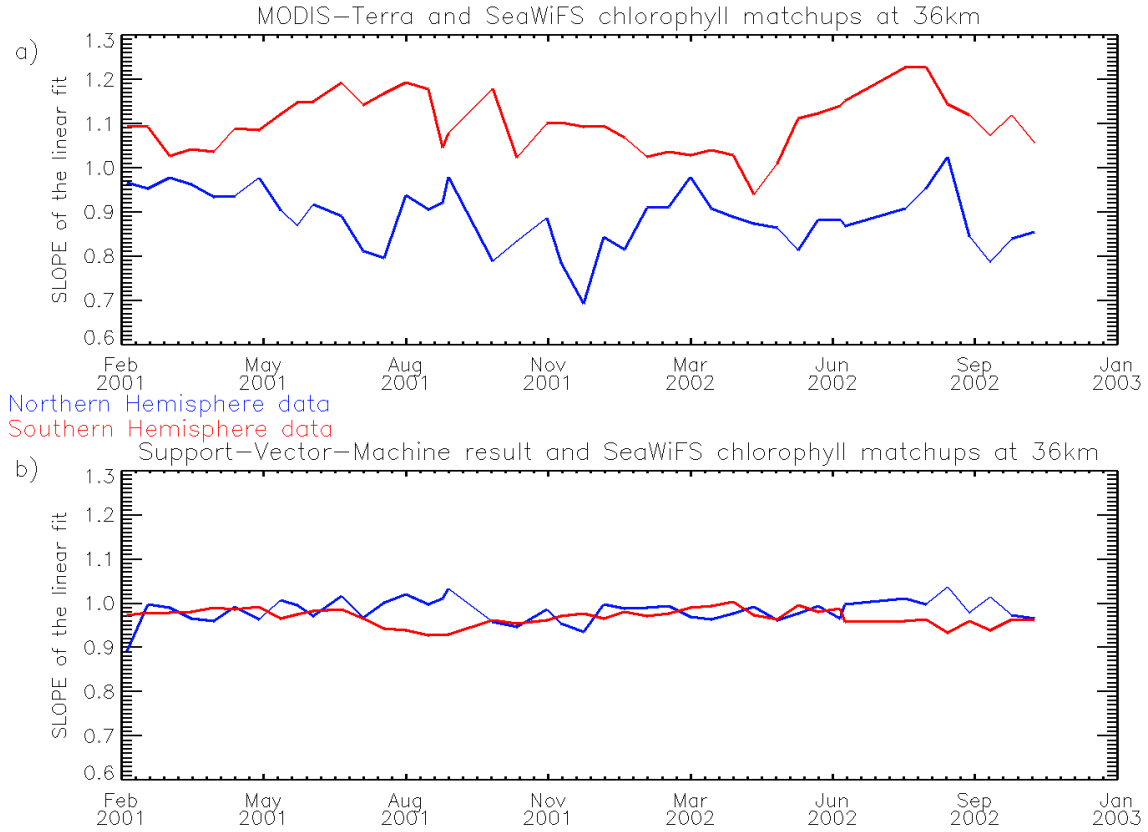



Fig. 5. Time trends in the slope of the linear fit between a) SeaWiFS and original MODIS chlorophyll for separate northern and southern hemisphere coverages and b) between SeaWiFS chlorophyll and the result of the support vector machine regression from MODIS data for the corresponding hemispheres.

MODIS and SeaWiFS ocean color comparisons from Chapter 2 demonstrated that discrepancies between the two sensor data manifested significant temporal and spatial variabilities and involved many dependencies among data variables. It was therefore challenging to extrapolate the knowledge of MODIS and SeaWiFS data relationships gained over a sparse time series onto the entire two-year period of the concurrent sensor coverage. There were two additional days in the data set which were not applied in the support vector machine training. Their purpose was to serve as a supplementary testing set to verify the support vector machine capabilities to extrapolate their knowledge through time to unknown dates. The days were inside the October 2000 to July 2002 period used in the training. Fig. 6 shows the scatter plots between SeaWiFS and original MODIS chlorophyll for these two dates, a), and between SeaWiFS chlorophyll and the cross-calibration result, b), where the cross-calibration support vectors were obtained from the 42-day training set. Fig. 6b illustrates that the cross-calibration knowledge gathered within the 42 days of combined MODIS and SeaWiFS coverage transferred

relatively well to the new data dates. The bulk of the support vector machine testing result closely approximated SeaWiFS chlorophyll. This was evident from improved scatter plot distributions. The MPD in plot b) was not as low as the value obtained on the 42-day testing set, which showed an average of 16.7% in Fig. 2, but it decreased from the original MODIS and SeaWiFS MPD in plot a). All other statistics were comparable or better than those achieved on the 42-day testing set, including a substantial improvement in the slope of the linear fit between SeaWiFS and the cross-calibrated chlorophyll.

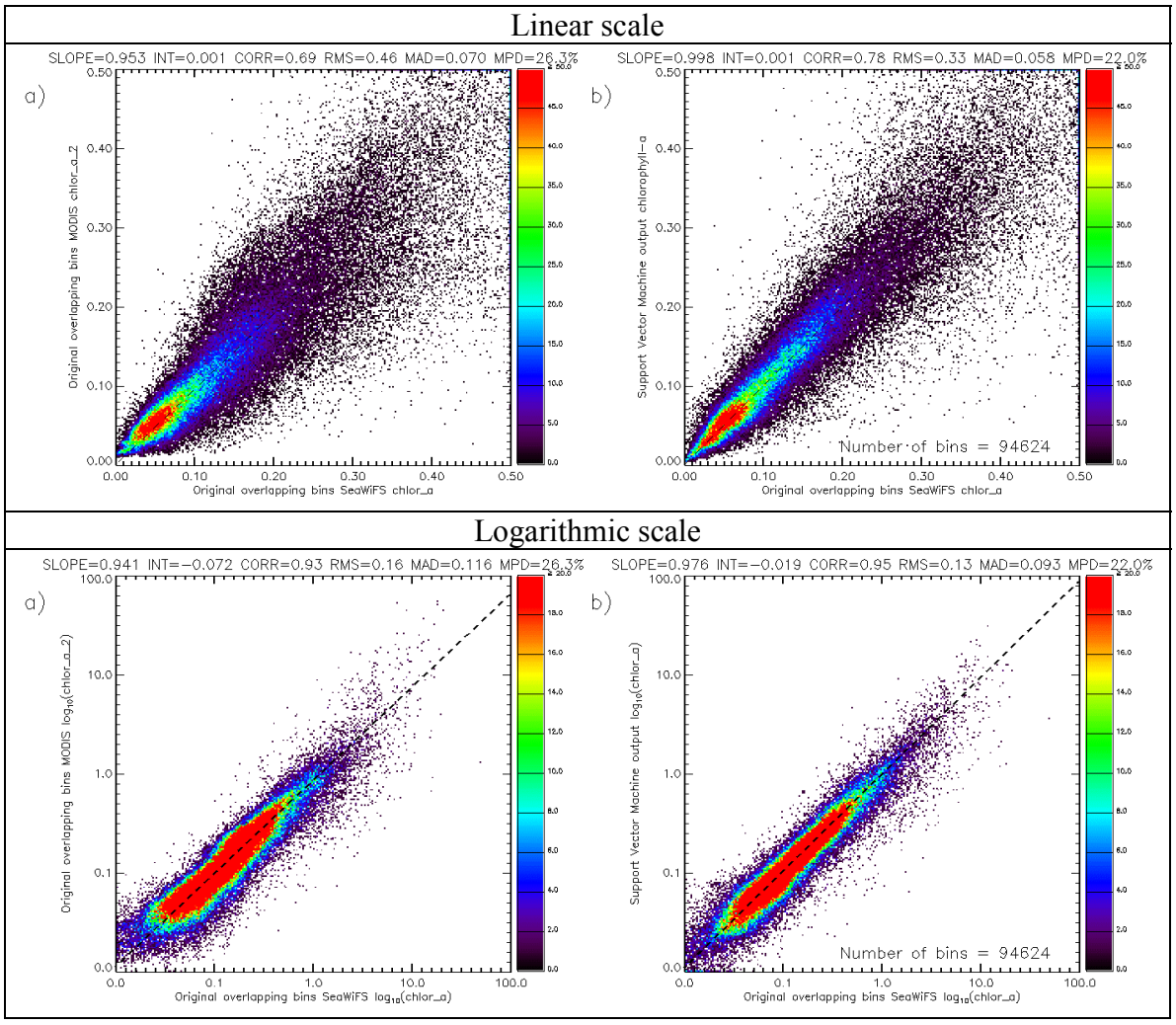



Fig. 6. Scatter plots for the 2-day testing data set separate from the 42-day time series used in the support vector machine training: a) SeaWiFS versus MODIS testing set chlorophyll, and b) SeaWiFS testing set chlorophyll versus chlorophyll regressed from MODIS data. The regression was able to extrapolate its cross-calibration knowledge through time to new data days and substantially improve the distribution of the bulk of MODIS chlorophyll data compared to SeaWiFS chlorophyll.

The cross-calibration was also investigated for its ability to extrapolate its knowledge of scan angle and latitudinal dependencies in MODIS data onto unknown data dates. The 2-day testing set was subsampled into western and eastern MODIS scan edge coverages

and northern and southern hemispheres. The results of the corresponding matchups are displayed in Fig. 7. The figure demonstrates that the support vector machines were still effective at eliminating scan angle dependencies and northern-to-southern hemisphere discrepancies in MODIS data. Consequently, their cross-calibration experience gained with a limited time series could be transferred to the complete time span of data.

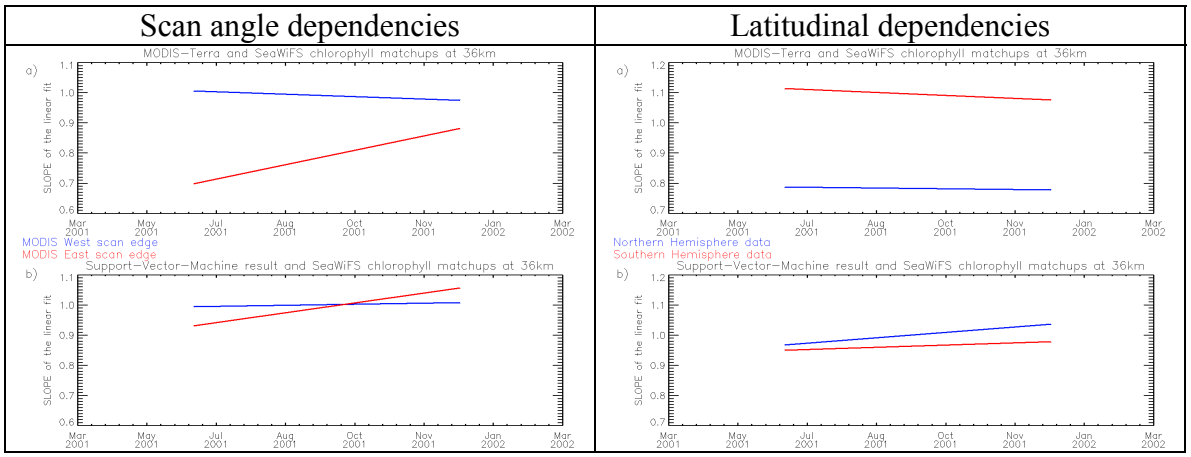



Fig. 7. Time trends in the slope of the linear fit for data corresponding to separate western and eastern parts of the MODIS scan and northern and southern hemisphere coverages between a) SeaWiFS and original MODIS chlorophyll and b) between SeaWiFS chlorophyll and the result of the support vector machine regression from MODIS data.

The goal of sensor cross-calibration to reproduce uniform SeaWiFS baseline response from MODIS data was consequently accomplished. Spatial and temporal discrepancies between MODIS and SeaWiFS data and scan angle and latitudinal dependencies in MODIS data were significantly reduced. Support vector machines were also able to extend their knowledge of complex relationships between MODIS and SeaWiFS ocean data to new unknown cases. A consistent series of daily global chlorophyll measurements could then be produced from MODIS and SeaWiFS by using MODIS data cross-calibrated with SeaWiFS. Fig. 8 contains merged MODIS and SeaWiFS global chlorophyll for 14 May 2001, where MODIS unique coverage for this day was regressed to the SeaWiFS baseline.

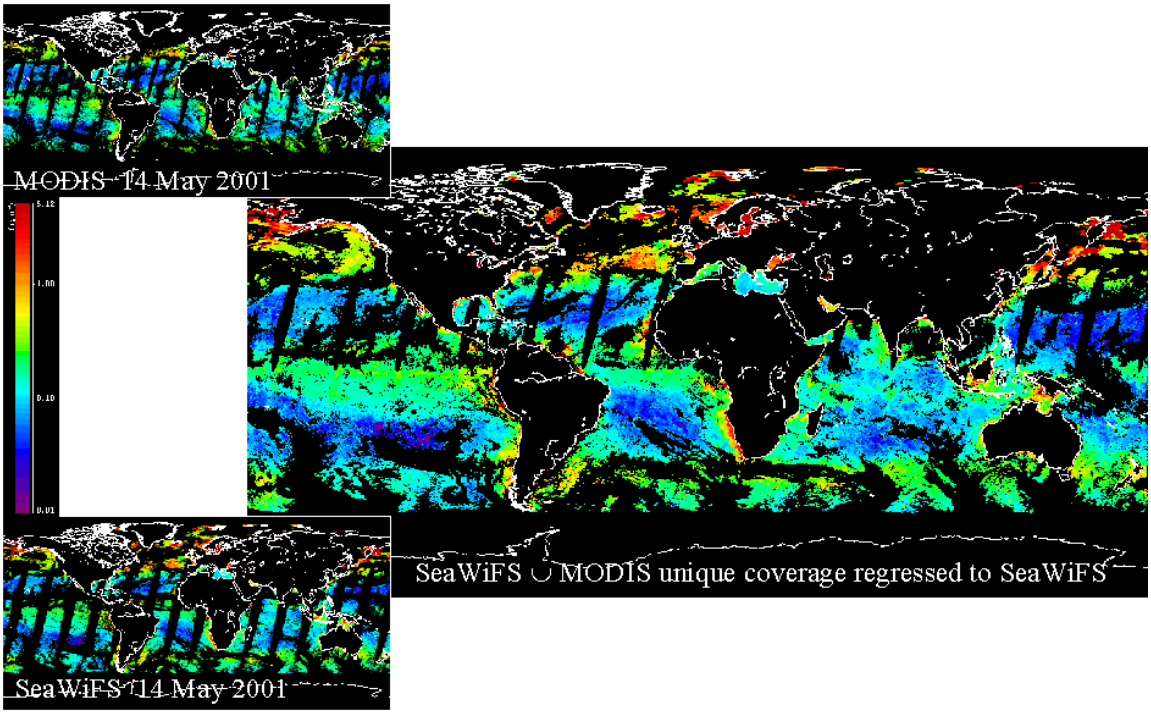



Fig. 8. Merged MODIS and SeaWiFS daily global chlorophyll baseline.

*1.4 Conclusions*

The objective of creating global ocean color data sets from multiple satellite sensors is important in the era of many concurrent ocean-observing missions. This goal is, however, hampered by incompatibilities in product data between the missions. MODIS-Terra and SeaWiFS ocean color data sets revealed significant discrepancies, described in Chapter 2, which were dependent on sensor calibrations and operational characteristics. These discrepancies inhibited the creation of consistent daily global merged data sets from both sensors. To bring MODIS ocean data to the SeaWiFS baseline, the application of machine learning cross-calibration was investigated. Support vector machines were trained to emulate SeaWiFS baseline chlorophyll from MODIS data. The ultimate objective was to produce joint MODIS and SeaWiFS daily global coverages which had the accuracy and the spatial and temporal consistency of SeaWiFS data sets.

Machine learning regression turned out to be a promising tool for the data merger. Support vector machines were able to accurately learn complex relationships between MODIS and SeaWiFS data and to effectively reduce sensor data discrepancies and eliminate MODIS artifacts, such as seasonal trends, scan angle dependencies, and spatial variation. Overall, the machines performed well within the time series on which they were trained and also proved the capability to extrapolate their knowledge to the entire time span of concurrent operations of the instruments. The performance of the machines can be improved by forming training sets that are more representative of the total MODIS and SeaWiFS time series and by reducing the noise in the data. Also, the support vector

machine regression can be further investigated for parameters and implementation additions to make it more robust and accurate.

Although the machine learning approach presented in this paper regards cross-calibration of MODIS and SeaWiFS global chlorophyll-*a* products, the mapping into other sensor product data can also be performed. For example, MODIS data can be used to predict SeaWiFS nLw measurements at various wavelengths. Chlorophyll concentration can then be calculated from these radiances using the standard SeaWiFS OC4v4 algorithm (O'Reilly et al., 1998). Top of the atmosphere reflectances can be mapped between the instruments given different sensor and solar geometries and atmospheric paths. Also, various other sensor parameters, such as adjustments to sensor calibration gains, can be mapped using the machine learning methodology.

## 2. Statistical Objective Analysis: Spatial and Temporal Interpolation of Multi-Sensor Ocean Color Data onto Daily Global Binned Coverage

### *2.1 Introduction*

Spatial and temporal interpolation can be applied to merge multi-sensor daily ocean color data onto global coverage grids (Kwiatkowska and Fargion, 2002b). The interpolation can be used in two ways. Firstly, single sensor data can be adeptly interpolated onto a global grid and to fill sensor's gaps in ocean coverage. Multi-sensor data can then be integrated by joint binning, a technique that is comparable to the "big-bin" approach or by optical algorithms using combined nLw retrievals at sensor-specific wavelengths (Maritorena *et al.*, 2002; Maritorena *et al.*, 2000). This would avoid the circumstances which limit the spatial consistency of the merged coverage, when some bins contain data from only one of the sensors and some bins contain a mix of data from multiple sensors. Secondly, multi-sensor data can be concurrently integrated onto a global grid using corresponding sensor data accuracies. Both methods of interpolation should preferably work with temporally and spatially stable ocean color measurements because applying time and space dependent instrument error estimates may be ambiguous and impractical.

For the purpose of this ocean color merger application, spatial and temporal interpolation was envisioned as a binning approach for multi-sensor data. The binning could fill many gaps present in daily global ocean color coverage depending on space and time distances between gaps and existing data. The binning was designed to operate on ocean color output products, such as chlorophyll-*a* concentration. Interpolation considered a spatial and temporal correlation structure of the chlorophyll field, which was dependent on the local area natural variabilities. Prior sensor cross-calibration was performed, such as in Section 1, to bring the multi-sensor data to a consistent baseline and eliminate sensor temporal trends and data artifacts. Sensor data accuracy used as a weighting factor in the interpolation was calculated from matchups with *in situ* measurements.

A well-known method to perform interpolation of environmental data is statistical objective analysis (Thiebaux and Pedder, 1987). Objective analyses use numerical methods to estimate geophysical field variables on surfaces and on three- and four-dimensional grids from data that are available at discrete locations and times. The method was first applied in meteorology to ground and satellite measurements. It calculates an interpolated grid-point value as a weighted linear combination of observations (Thiebaux, 1973). In an empirical linear interpolation, weights are either a function of separation between analysis and observing locations or a function of accuracy of one observation relative to another. A distance weighting with weight normalization is the most common. In statistical objective analysis, the approximation is obtained by making additional use of the ensemble spatial correlation structure of the whole field, i.e. the spatial distribution of observations relative to one another (Julian and Thiebaux, 1975). The analysis considers instrument errors and other variations in data so that an interpolated value of a field variable does not have to be identical with an observed value at corresponding space/time coordinates, but it is intended to coincide with the signal component of the observed variable.

The practical requirement for the use of this algorithm is that there has to exist a preliminary "prediction" of the signal, or a first-guess field, and the objective analysis corrects this prediction by interpolating the signal with a single or multiple passes of the algorithm. At successive corrections, non-zero weights are given to observed increments only if the observations lie within a prescribed distance, known as the influence radius, of the grid point being considered. This influence radius may be decreased with successive passes of the algorithm. The analyzed grid point value is written as

$\boldsymbol{Y}_o = \sum_{j=1}^{m} \boldsymbol{h}_j (\boldsymbol{X}_j^o - \boldsymbol{X}_j^f) + \boldsymbol{X}_o^f$, where m is a number of observing locations, $Y_o$ is the interpolated value of a grid point, $X^o_j$ is an observation value at point $j$, $X^f_j$ is a first guess for the $j$th observation point, $X^f_o$ is the first guess for the analyzed grid point, and $\boldsymbol{h}$ is the weight vector. The weights are obtained by minimizing the ensemble average of the squared difference between the analysis value and the true value of the field signal (Thiebaux and Pedder, 1987). The solution to this minimization problem is a covariance array for the joint distribution: $\begin{pmatrix} \boldsymbol{h}_1 \\ \vdots \\ \boldsymbol{h}_m \end{pmatrix} = \begin{pmatrix} \boldsymbol{\sigma}_{11} & \cdots & \boldsymbol{\sigma}_{1m} \\ \vdots & & \vdots \\ \boldsymbol{\sigma}_{m1} & \cdots & \boldsymbol{\sigma}_{mm} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{\sigma}_{01} \\ \vdots \\ \boldsymbol{\sigma}_{om} \end{pmatrix}$, where $\sigma_{ij}$ is a covariance between an $i$th and $j$th observation.

Statistical objective analysis has been applied to create NOAA's real-time global sea surface temperature (SST) maps (Reynolds, 1988; Reynolds and Smith, 1994). The maps are produced weekly on a one-degree grid. The analysis uses buoy, ship, and satellite SST data, and SST's simulated by sea-ice coverage. The approach applies individual sensor errors and a globally averaged space-lag correlation structure of the SST field.

*2.2 Statistical Objective Analysis Implementation*

For this study, the covariances between chlorophyll data points were expressed in terms of space and time-lag correlation functions, ρ(s), which were calculated from ocean color data. This assumed that the variance of the chlorophyll concentration truth-value was a constant $\sigma^2$ at all locations, the noise variance of chlorophyll was a constant $\eta^2$ and independent of location, the space-lag covariance of the chlorophyll-truth was isotropic, and the covariances of the noise at different locations were zero (Thiebaux and Pedder, 1987). The weighting scheme for the chlorophyll-interpolated truth-value was then written as: $\begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ h_m \end{pmatrix} = \begin{pmatrix} 1+\gamma & \rho(s_{12}) & \cdots & \rho(s_{1m}) \\ \rho(s_{12}) & 1+\gamma & \cdots & \rho(s_{2m}) \\ \vdots & & \ddots & \vdots \\ \rho(s_{1m}) & \rho(s_{2m}) & \cdots & 1+\gamma \end{pmatrix}^{-1} \begin{pmatrix} \rho(s_{01}) \\ \rho(s_{02}) \\ \vdots \\ \rho(s_{0m}) \end{pmatrix}$, where $s_{ij}$ is a distance between the *i*th and *j*th points and $\gamma^{-1}$ is the signal-to-noise ratio, $\sigma^2 / \eta^2$, of the chlorophyll concentration product.

The assumptions for this equation were not met for chlorophyll data and modifications were introduced into the algorithm. Chlorophyll truth and noise variances vary depending on the amount of chlorophyll concentration which fluctuates through five scales of magnitude from 0.001mg/m$^3$ to 100mg/m$^3$. Because chlorophyll has a lognormal distribution, analyzing chlorophyll in a logarithmic space shrinks the range of data values to a single scale of magnitude (Campbell, 1995). To use this property, chlorophyll and chlorophyll-error variances were derived from *in situ* matchups in the logarithmic chlorophyll space and the logarithmic signal-to-noise ratio was assumed constant at all locations. For each interpolation grid point, chlorophyll values were then converted to the logarithmic space to tie in with the logarithmic statistics.

Modeling space-lag correlation functions is a subject of extensive research (Julian and Thiebaux, 1975). The correlation is expressed as a function of the spatial separation of locations of points in geographic coordinates. For this study, a 3-dimensional statistical objective analysis was investigated using time as the third dimension and with data points separated by a day or a number of days to interpolate a given grid location. For the 3-dimensional analysis, space-lag correlations were derived for data at different distances and 0 to 7 days apart. The space and time-lag correlation functions were initially calculated globally over daily chlorophyll concentration fields. Afterward, non-isotropic space and time-lag covariance of the chlorophyll-truth was investigated. Space and time-lag correlation of the chlorophyll field was assumed to be dependent on local area spatial and temporal variabilities. Chlorophyll variabilities were modeled using a standard deviation function. The variabilites were derived using 9-day and biweekly MODIS-Terra and SeaWiFS global L3 chlorophyll maps at an initial spatial resolution of 36km, chosen to limit processing time. Standard deviations were dependent on the radius of the local area under investigation and on the average chlorophyll magnitude within the area. Standard deviations were approximated for different chlorophyll magnitudes and 13 classes of ocean variabilities were defined based on the standard deviation functions. Each ocean data point, a bin, on the global map was then assigned to a vector of

chlorophyll variability classes. The vector was composed of chlorophyll variability classes at consecutive distance ranges from 0 to 1000km at 10km intervals from the point under consideration. Space and time-lag correlations could then be calculated in a manner dependent on the classes of chlorophyll variability in the ocean. Correlations were obtained for chlorophyll value increments from the first-guess background field at different distance ranges. Separate increment correlations were calculated spatially within a single given day and between the given-day chlorophyll and chlorophyll a number of days away. The background field was assumed to be a global chlorophyll 9-day mean. To compute the correlations, data were applied from days which followed the week used as the first-guess estimate. The correlations across consecutive distance ranges were calculated separately for each chlorophyll variability class. The components for the increment correlation functions came from the global chlorophyll maps and, for each point, used its variability classes at corresponding distance intervals. Fig. 9 shows preliminary results of space-lag correlations for increments from the 9-day global chlorophyll first-guess field at the same day and 2-day time intervals. The results of the calculations were approximated by exponential functions. The 13 correlations are shown in different colors corresponding to their classes of chlorophyll spatial variabilities, from low variance in dark blue to high variance in red. The figure also displays global ocean variability maps for the 13 spatial variability classes with the same color-coding. The first map shows classes of variability within a 100km radius and the second – classes of variability within a 600km radius.

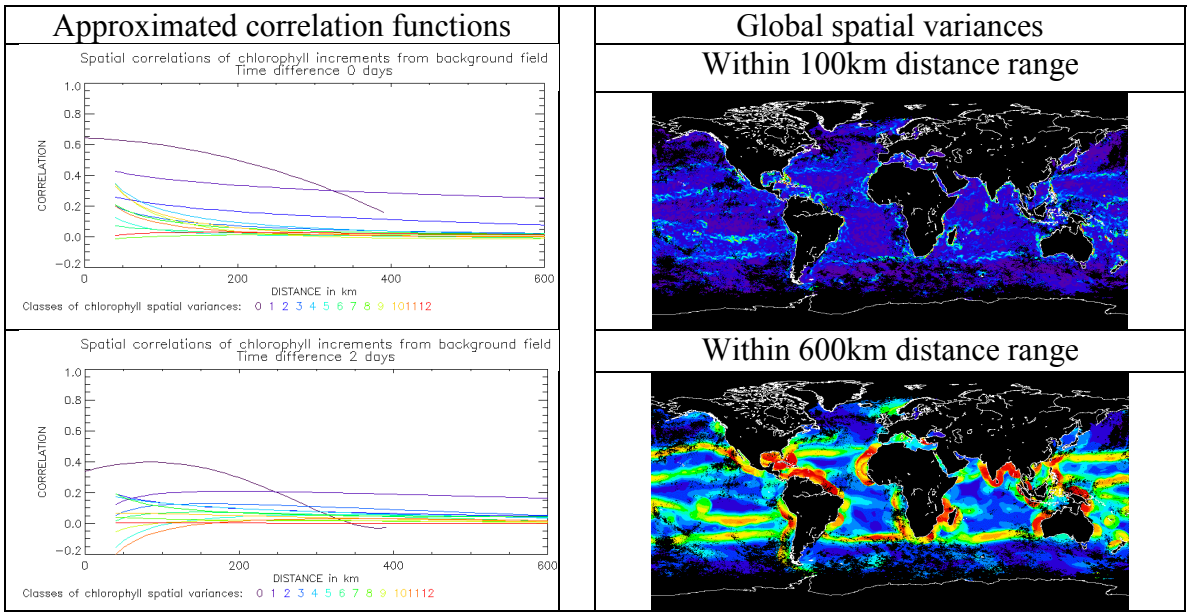



Fig. 9. Space-lag correlation functions for chlorophyll increments from the first-guess 9-day mean field averaged over global data sets and within 10km distance intervals between the points. The 13 functions shown in different colors correspond to different spatial variability classes, from low variance in dark blue to high variance in red. Global spatial variability distributions are shown on the right where the variances where calculated within 100km and 600km distance ranges, correspondingly. The correlations were approximated from MODIS-Terra L3b 36km global time series.

Fig. 9 illustrates that natural spatial variability of phytoplankton limits the extent of spatial and temporal interpolation of ocean color data. The spatial-lag correlations for chlorophyll increments were relatively low. The class number 1 at time difference of 0 days had correlation values around 0.4 for short distances which did not decrease to 0 for longer distances. High variability classes formed increment correlation functions not higher than 0.4 and approximating 0 at short distances. This apparent spatial diversity of chlorophyll concentration increments, averaged over global scales, showed that chlorophyll data differed from meteorological data, which were better correlated and for which the statistical objective analysis was originally created (Thiebaux and Pedder, 1987). The correlation functions were dependent on the definition of the first-guess background field. A relatively up-to-date and complete chlorophyll map had to be chosen to initialize the analysis. Eventually, the first-guess field would be the previous day global chlorophyll coverage obtained by the preceding step of the interpolation. Therefore, the correlation functions could change somewhat when a more appropriate first-guess field is applied. The influence radius for the analysis was defined to be equal to 600km, following information about the shape of the space-lag correlation functions for chlorophyll increments. With low correlation values and the short influence radii, the statistical objective analysis cannot interpolate chlorophyll grid points which lie relatively far from valid data.

*2.2 Statistical Objective Analysis Results*

Because the current investigations of the statistical objective analysis were preliminary, only single-sensor-based interpolation was performed to fill sensor's gaps in ocean coverage (Kwiatkowska and Fargion, 2002b). A simple logarithmic signal-to-noise ratio of chlorophyll data was calculated by dividing the mean value of chlorophyll by the chlorophyll variance both derived from the matchups with *in situ* measurements. Two-dimensional and 3-dimensional forms of the analysis were tested only for globally averaged chlorophyll increment correlations without considering local area variability ranges. A globally averaged space-lag correlation function for chlorophyll-*a* concentration in single-day increments is illustrated in Fig. 10. The function was created from chlorophyll-increment correlations from a weekly mean and was approximated using an exponential function: $\mathbf{y} = 0.582708 \cdot 0.984943^{\mathbf{x}} + 0.0189976$, where X was a distance between points expressed in kilometers. The function was obtained using SeaWiFS L3b daily global chlorophyll time series at 36km resolution. A value of 1 was assumed when points within the same-day overlapped spatially and a value of 0 was assumed when the points were outside the radius of influence, which was set at 600km. Space-lag correlation functions were similarly approximated for points separated by up to 3 days from the interpolated day.

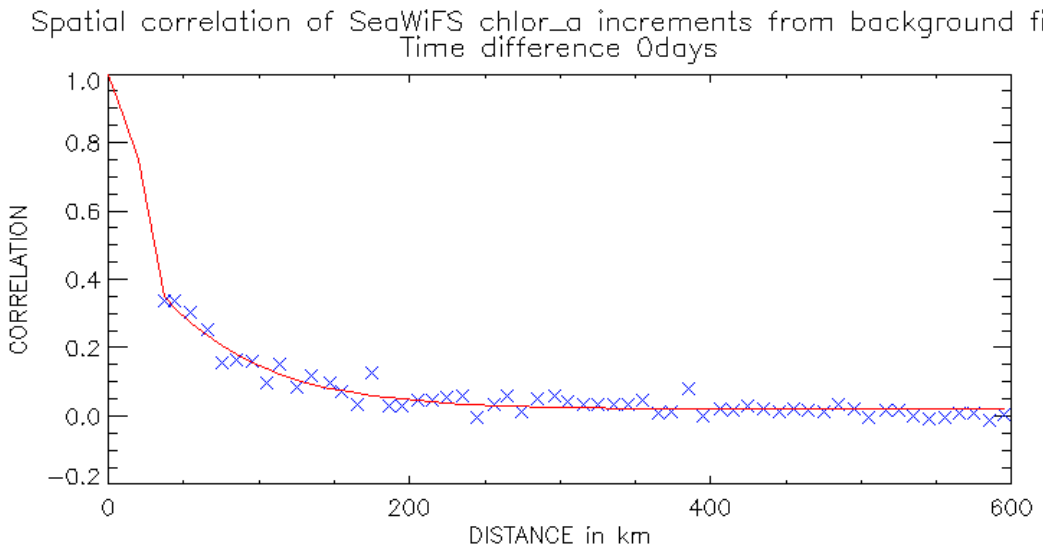



Fig. 10. Chlorophyll increment space-lag correlation function averaged over a single-day global data set and within 10km distance intervals between the points. The first-guess background field was the weekly global chlorophyll mean.

Fig. 11 shows the result of the spatial statistical objective analysis on SeaWiFS L3 binned daily global chlorophyll coverage at 36km resolution for 8 April 2001. Only those SeaWiFS grid points were interpolated which, for this day, coincided in coverage with MODIS bins containing valid data. If a similar interpolation was done using this day's MODIS data, MODIS and SeaWiFS chlorophyll concentration products could be merged by means of averaging or weighted averaging, where all valid bins would contain data from both sensors. For the analysis to be applied with the semi-analytical optical algorithm (Maritorena *et al.*, 2002), the interpolation had to be done on MODIS and SeaWiFS nLw products on the bands used for chlorophyll extraction. Ultimately, the interpolation would be performed jointly on MODIS and SeaWiFS chlorophyll using corresponding statistics for both sensors and, preferably, as a binning scheme beginning with the L2 products. MODIS data would then be first cross-calibrated with SeaWiFS because, otherwise, MODIS trends and artifacts would be inseparably intertwined with SeaWiFS data in the merged data set.

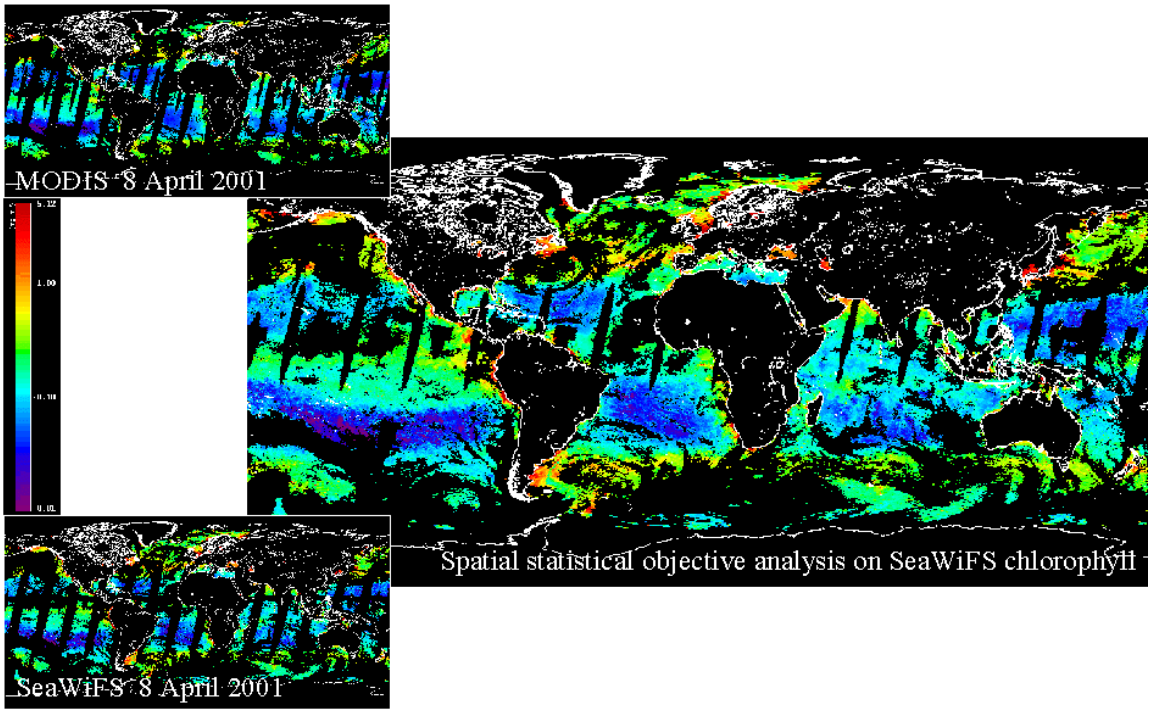



Fig. 11. Original MODIS and SeaWiFS 36km binned chlorophyll concentration data sets for 8 April 2001 and the result of the 2-dimensional statistical objective analysis on the SeaWiFS chlorophyll bins coinciding with the MODIS coverage.

For the interpolation, problematic areas in ocean color daily imagery were those where gaps in global coverage were large in spatial and temporal terms, such as below persistent clouds, sun-glint, or SeaWiFS tilt change. It was observed that within these gaps the result of the analysis looked realistic but some interpolated coverage had values very close to the first-guess field. Eventually, each output point would be assigned a confidence level which would be dependent on accuracies of the original data sets and on the distances from the existing points used for the interpolation.

The statistical objective analysis was computationally involved. It considered ensemble spatial distributions of observations relative to one another which were contained within a radius of influence of an investigated grid point. This resulted in large covariance matrices whose size depended on the number of valid data points within the radius of influence from the interpolated point. At 9km, which is the ultimate resolution for the data merger, the quantity of L3 bins will be an order of magnitude higher than at the resolution of 36km. When ultimately operating on L2 data, the amounts of points analyzed inside the radius of influence could be massive. The analysis involved inverting square matrices of covariances for each investigated grid point. To ease the computational effort, an effective strategy was designed in which radii smaller than the influence radius were first searched to determine whether they contained sufficiently high proportions of valid data bins to perform the interpolation. If there were enough valid data inside the smaller radius, these data were then used to interpolate the grid point.

Because the covariance matrices were symmetric and positive definite, an efficient Cholesky decomposition was used to solve the matrix equation (Press *et al.*, 1992).

### *2.3 Conclusions*

Statistical objective analysis was introduced as a spatial and temporal interpolation approach to combine multi-sensor ocean color data sets onto daily global grids using corresponding sensor accuracies and an ensemble correlation structure of the global chlorophyll field. The interpolation was envisioned as a binning approach for multi-sensor data beginning with the L2 products. The binning would effectively combine sensor data using pixels surrounding the bin grid points in space and across time and would consequently fill many gaps in daily global ocean color coverages. The ensemble correlation structure of the chlorophyll field was established individually for all global coverage grid points and made dependent on the local area natural variabilities. Preceding the interpolation, sensor cross-calibrations would be performed to bring the multi-sensor data to a consistent baseline and eliminate sensor temporal trends and data artifacts. Sensor data accuracy was used as a weighting factor in the interpolation and came from matchups with *in situ* measurements.

The initial results were obtained from the 2- and 3-dimensional statistical objective analysis of daily global SeaWiFS L3 binned chlorophyll data. The analysis interpolated selected missing SeaWiFS bin coverage for this day. The analysis demonstrated to be a useful tool for ocean color data merger. However, more research would be needed to make the statistical objective analysis more effective in terms of the choice of the first-guess background fields and associated space-lag correlation functions and influence radii. The statistical objective analysis was also computationally involved in operational processing of multi-sensor data. Therefore, means for improvement of its efficiency could be investigated. Finally, the capabilities of the analysis to provide error bars for all interpolated data points could be further studied.

## 3. Local Area Application of Data Merger: Enhancement of Oceanic Features in Lower Resolution Imagery Using Higher Resolution Data

### *3.1 Introduction*

This study examined ocean color merger opportunities at local spatial scales to provide useful tools for scientists interested in smaller-size geophysical phenomena and in complex environments such as coastal zones. The feasibility of merging ocean color data from sensors of different spatial resolutions was studied for cases where there was overlapping ground coverage for individual scenes (Kwiatkowska-Ainsworth, 2001; Kwiatkowska and Fargion, 2002a). The prospect of enhancing oceanic features in lower resolution imagery through the use of higher resolution data was also investigated.

The algorithm operated on L2 ocean color data products and was based on a signal processing approach — wavelet multiresolution analysis (Rioul and Vetterli, 1991). The

wavelet transform enabled an image to be examined at different frequency and scale intervals (Mallat, 1989). This corresponded to image analyses at variable frequency and spatial resolutions. The resolution of an image, corresponding to a measure of detail information in the scene, was defined and changed by a combination of high pass and low pass filtering operations. The scale of an image was altered by downsampling and upsampling operations. The wavelet transforms used in this analysis therefore functioned as power-of-2 operators for subsampling and resolution change. Fig. 12 illustrates the process of decomposition of a one-dimensional signal by a discrete wavelet transform (DWT). The figure also shows changes in scale and frequency contents of the filtered output.

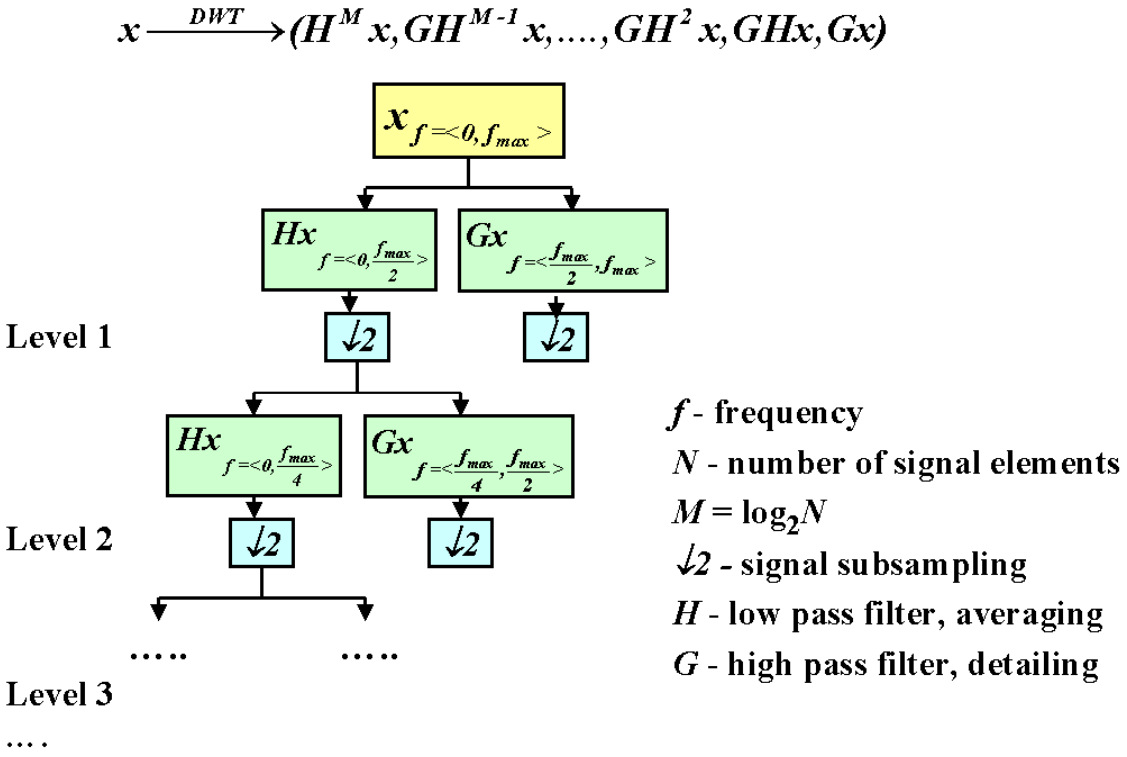



Fig. 12. Consecutive levels of wavelet transform decomposition of a one-dimensional signal. At the first decomposition level, the signal is passed through the high pass and low pass filters, followed by subsampling by 2. The output of the low pass filter is then passed at level 2 through the same low pass and high pass filters for further decomposition. The process is repeated at the subsequent levels. Changes in output frequency and scale are indicated.

The wavelet merger algorithm thus operated on scenes from sensors of different spatial resolutions (Núñez *et al.*, 1999; Blanc *et al.*, 1998). The high-frequency, low-scale spatial detail in the higher resolution scene was extracted using the high pass filters of the wavelet transform. The result of the low pass filtering of the higher resolution image was completely replaced by the lower resolution scene. This modified wavelet transform of the higher-resolution image was then reversed. For the lower resolution scene this process resulted in the increased spatial resolution and added high frequency variation.

The enhanced spatial resolution was gained without altering the mean magnitudes of lower-resolution ocean color values. This made the wavelet method particularly useful when the quality of data was different between the sensors and the measurement accuracy of the lower resolution sensor had to be preserved. To perform merger of data from both scenes, the result of the low pass filtering of the higher resolution image, instead of being completely substituted by the lower resolution scene, would be replaced with its weighted average with the lower resolution scene. The reversal of the transform would then produce a merged image where the merger was performed on the level corresponding to the lower-resolution coverage from both sensors and the lower scale detail was added from the higher resolution scene.

*3.2 Wavelet Transform Implementation and Results*

The wavelet algorithm was tested using chlorophyll-*a* concentration imagery from SeaWiFS and MOS. The SIMBIOS Project cross-calibrated SeaWiFS and MOS missions, processed their data uniformly, and analyzed them for overlapping concurrent ground-coverage (Wang and Franz, 2000). SeaWiFS L2 HRPT and LAC scenes used in the analysis had a native resolution of 1.1km and MOS imagery had the resolution of 0.5km. A significant obstacle was to co-register scenes from both sensors so that the accuracy of the overlay was within 0.5km and to define resolutions and scene sizes to be in power of 2 for the multi-resolution analysis. A basic strategy was designed where SeaWiFS data were binned at 1km and MOS data were binned at 0.5km. Bins were then projected onto a rectangular longitude/latitude grid map to facilitate image processing. Because the dimensional sizes of the rectangular grids were limited to powers of 2, the mapped scenes had to be padded with zeros to fill the grid. The size of the SeaWiFS grid was half the size of the MOS grid. To preserve the spatial resolution of the bins, the projection spread the bins longitudinally according to the longest row for each scene. Bins from all other rows were then fitted into the grid given their longitude distance from the longest-row longitudes. This technique was only applicable to local coverage scenes used in this analysis, which were of the LAC and HRPT size, and it would not be appropriate for global imagery. Any missing grid points caused by the mapping of spherical coordinates onto a rectangular grid were approximated. The approximation used a wavelet-based iterative algorithm that minimized high frequency anomalies associated with the missing data points. The preprocessing therefore resulted in reformatted mapped scenes from both sensors with the desired size and resolution for the wavelet analysis.

Because the resolution ratio between SeaWiFS and MOS scenes was equal to 2, only one level of wavelet decomposition was required for the processing. This single pass of the wavelet filtering was applied to the MOS image. The transform extracted pixel-to-pixel spatial detail from MOS data in its high-frequency components and higher-scale background in its low-frequency components. The transform also subsampled the MOS scene by 2 so that its high and low frequency components corresponded to 1km spatial resolution, the same as the SeaWiFS scene resolution. The SeaWiFS image was concurrently preprocessed to bring the magnitude of its chlorophyll values to the level corresponding to a single application of the low pass filter. Then in the two scenarios, the MOS low-pass filter result was replaced either by the entire preprocessed SeaWiFS scene

or by its weighted ratio with the preprocessed SeaWiFS scene. The ratio depended on the established relative accuracies of the chlorophyll products from each instrument. An inverse wavelet transform was subsequently applied which produced an increased 0.5km resolution SeaWiFS image or a 0.5km SeaWiFS image merged with MOS data. The enhanced SeaWiFS scene inherited its low-scale spatial detail from the high-frequency contents of the MOS scene. To generate the final product, flags and masks from SeaWiFS and MOS chlorophyll data were also merged and applied to the subsequent image. Fig. 13 shows an example of the wavelet multiresolution merger of SeaWiFS and MOS scenes of Mallorca and Menorca in the Mediterranean Sea.

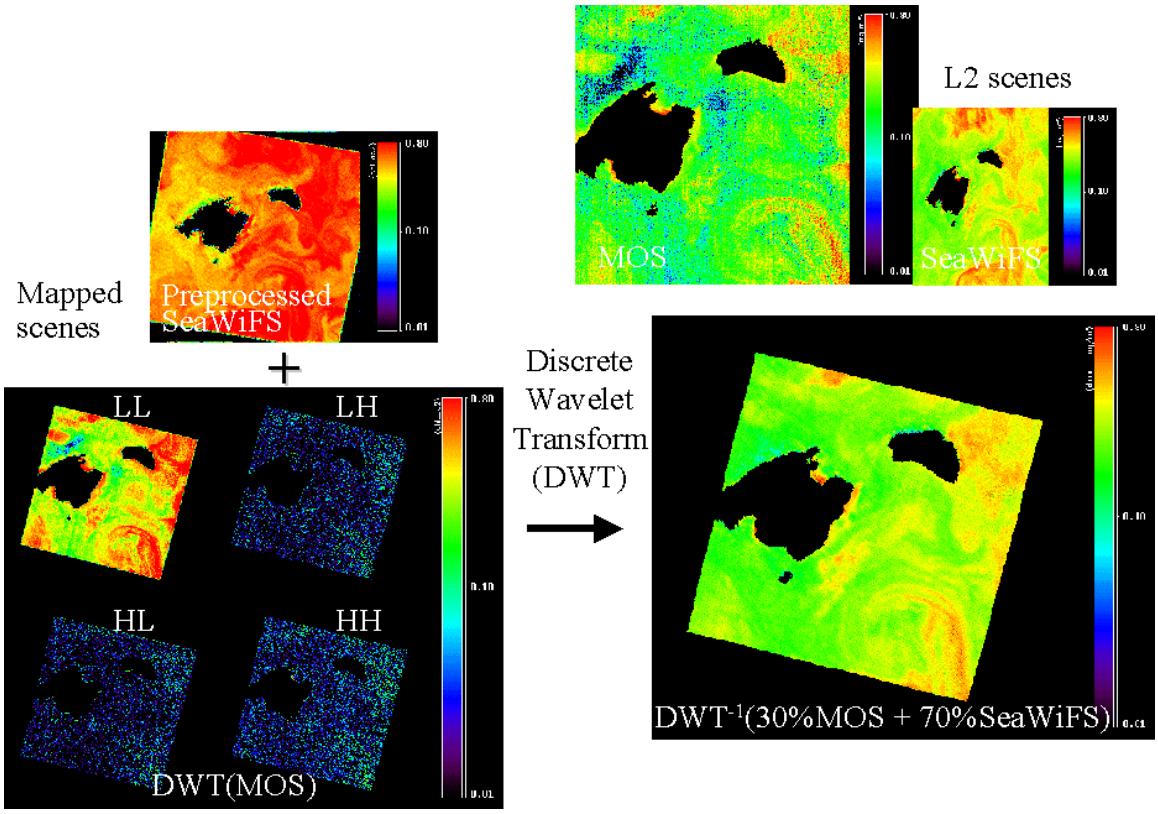



Fig. 13. Original L2 MOS and SeaWiFS chlorophyll-*a* concentration scenes and the process of wavelet-based merger of both data sets mapped to a rectilinear grid. The SeaWiFS scene was preprocessed and the MOS scene underwent a single level of wavelet decomposition. The 2-dimensional DWT separately passed high (H) and low (L) pass filters through the scene at first across the rows with column subsampling, and then across the columns with row subsampling. The MOS row and column low-pass coefficients (LL) were replaced by their weighted ratio with the preprocessed SeaWiFS scene. The wavelet transform was reversed, thus producing the merged output image at 0.5-km resolution.

To validate the wavelet algorithm, the original MOS scenes were compared against the SeaWiFS scenes enhanced to 0.5km resolution using the wavelet method and using bilinear interpolation. The bilinear interpolation on its own did not provide the benefits of the higher-frequency feature extraction which enabled SeaWiFS imagery to acquire

spatial detail inherent in MOS data. Quantitatively, the correlation of bilinearly interpolated SeaWiFS imagery with original MOS imagery was considerably smaller (~10%) than the correlation for the wavelet-enhanced SeaWiFS scenes. Qualitatively, the gain in spatial detail obtained by the wavelet approach was consequential and unique. Merger of original MOS scenes with bilinearly interpolated 0.5km resolution SeaWiFS data was also compared against the result of the wavelet multiresolution analysis. This merger did not, however, allow the preservation of the magnitudes of SeaWiFS lower-resolution ocean color values and complete high-frequency spatial variation from MOS data. Overall, the wavelet algorithm performed superior to other approaches.

The application of the wavelet approach brought also some difficulties. MOS data were inherently noisy. Although the wavelet-merged scenes appeared sharper, there was a degree of high-frequency noise introduced from MOS scenes. As it happened, wavelets also provided a means for denoising speckled imagery (Donoho, 1995). Therefore, denoising was implemented as an option in the algorithm. The implementation was based on soft-thresholding of wavelet coefficients which was equivalent to removing Gaussian noise from an image. Additionally, manipulation of wavelet coefficients caused undesirable ringing effects in images because of the presence of high frequency features. To limit the ringing, a selected number of transformed solutions based on different wavelet functions was averaged. Daubechies_20, Coiflet, Haar, and spline functions were examples of the wavelet functions used.

*3.3 Conclusions*

This study examined possible applications of ocean color data merger at local spatial scales. It investigated integration of data from sensors of different spatial resolutions. It also determined the ability to generate merged products of the resolution equal to that of the higher resolution sensor. Inherent in the technique was an option to preserve in the merged output, mean high-scale ocean color values from the lower resolution coverage. This corresponded to an ability to enhance the lower-resolution ocean color baseline with low-scale spatial detail from the higher resolution data. Simultaneously, the baseline was able to retain its calibration quality. This would provide the useful tools for the investigation of smaller-size geophysical phenomena and complex environments, such as coastal zones. Wavelet-based multiresolution analysis was used to extract high-frequency low-scale features from high resolution imagery and transfer them to lower resolution scenes.

It would be of interest to apply the wavelet algorithm to the merger of overlapping scenes between MODIS and SeaWiFS GAC so that SeaWiFS imagery could be enhanced by the spatial detail contained in MODIS data. A useful application would also be to combine MODIS or SeaWiFS ocean color products at 1km or 4km resolution with high frequency spatial information contained in MODIS high-resolution bands, such as 500m and 250m.

# 4. Local Area Application of Data Merger: Merger of Satellite and *In Situ* Measurements

*4.1 Introduction*

An approach was developed to merge L2 ocean color data with *in situ* measurements. The major purpose was to provide a utility to demonstrate changes in remotely sensed chlorophyll or nLw range and distribution when collected *in situ* measurements were overlaid upon local area scenes. The algorithm was intended for use in local area applications to verify remotely sensed ocean color data and provide a change visualization tool. The merger of satellite and *in situ* data was dependent on the spatial and temporal correlation structure of the ocean color field, which was by itself, contingent upon local area spatial and temporal variabilities, as shown in Section 1.

Merger was based on the application of the wavelet transform which spatially extended *in situ* data point values onto corresponding areas in satellite scenes (Kwiatkowska and Fargion, 2002a; Mallat, 1989). These areas were defined by a radius of influence and depended on the geographical location of *in situ* measurements (Barnes, 1964). The radius of the area of influence was defined using local texture estimates, such as the spatial variability classes defined in Section 2. The more irregular the texture was around the *in situ* measurement point, the smaller the radius; the smoother the texture, the bigger the radius. The Hann window function was applied to scale the effects of the *in situ* data points away from the area centers (Press *et al.*, 1992). Ultimately, a space-lag correlation function for a given area spatial-variability class would be used. The degree of change introduced by *in situ* measurements onto ocean color satellite scenes also depended on the established relative accuracies assigned to *in situ* and satellite data.

The methodology behind the wavelet merger was the following: Because *in situ* measurements were screened for quality, they were assumed representative of generalized ocean color conditions within their area. *In situ* data points were typically intended not to affect the local area low-scale spatial variabilities and not to change shapes of ocean patterns within the scenes. Each *in situ* observation was therefore associated with a low-frequency background ocean-color value corresponding to its coverage point. The low frequency background was extracted by the low-pass wavelet filter. The original wavelet coefficients of each scene were then replaced with the coefficients updated with the *in situ* data point and the point's values scaled smoothly towards the edges of the area of influence. The magnitude of the correction also depended on the estimates of the relative accuracies of satellite and *in situ* measurements. The wavelet thus forced the resulting satellite pixels to be interpolations of *in situ* data only within the low-resolution representation of the scenes. The high frequency coefficients of the updated imagery were left unchanged to preserve the original high-resolution spatial variabilities within the areas of influence and to protect spatial structures in the scenes.

*4.2 Satellite and In Situ Merger Implementation and Results*

Merger of satellite and *in situ* chlorophyll-*a* concentration observations was analyzed using SeaWiFS and California Cooperative Oceanic Fisheries Investigation (CalCOFI) data for the years 1997 and 1998. From the experience with ocean color validation, it was known that there was a significant scarcity of contemporaneous satellite and *in situ* observations, mainly because of the presence of clouds, sun glint, coverage gaps between satellite orbits, and other satellite viewing and meteorological conditions. Merger was dependent on the time difference between the satellite overflight and *in situ* data collection. The maximum time span between SeaWiFS and *in situ* observations was set for 12 hours, although the ultimate time span would be dependent on the local area spatial variability and the corresponding time-lag correlation of the chlorophyll field. Within the 12-hour time difference, there were just 13 SeaWiFS L2 LAC and HRPT files with concurrent satellite and CalCOFI measurements for which the merger could be performed. One file out of the 13 contained three points within the scene. The low number of matchups was principally caused by the presence of cloud cover.

To limit the cases where small clouds (a few pixels long) and other conditions caused ocean color pixels to be masked out from the imagery, a gap-filling algorithm was implemented. Its goal was to preserve spatial patterns of chlorophyll distributions in ocean color scenes without smoothing. The algorithm was based on an iterative reduction of the total of high frequencies associated with missing pixels in the analyzed scene. The high frequency content of a pixel was established by inverting a wavelet transform output of the scene where the inversion was limited to the result of the forward transform high-pass filter. The iterations were initialized by filling the gaps with values corresponding to the lower frequency representation of the scene. To eliminate local minima, a random perturbation was introduced to the best values for the gap pixels which were found by the recursive search. The gap-filling approach was implemented in combination with the satellite and *in situ* measurement merger to eliminate small clouds within the areas of influence of *in situ* data points. This produced an increase in matchups of about 10%.

A sequential processing algorithm was implemented for all *in situ* data points extracted from selected SeaBASS records (Werdell and Bailey, 2002) and a corresponding list of SeaWiFS L2 files. The algorithm processed image subscenes encompassing areas of influence of consecutive *in situ* points and fused the points into the images. Examples of the *in situ* and satellite data merger are displayed in Fig. 14.

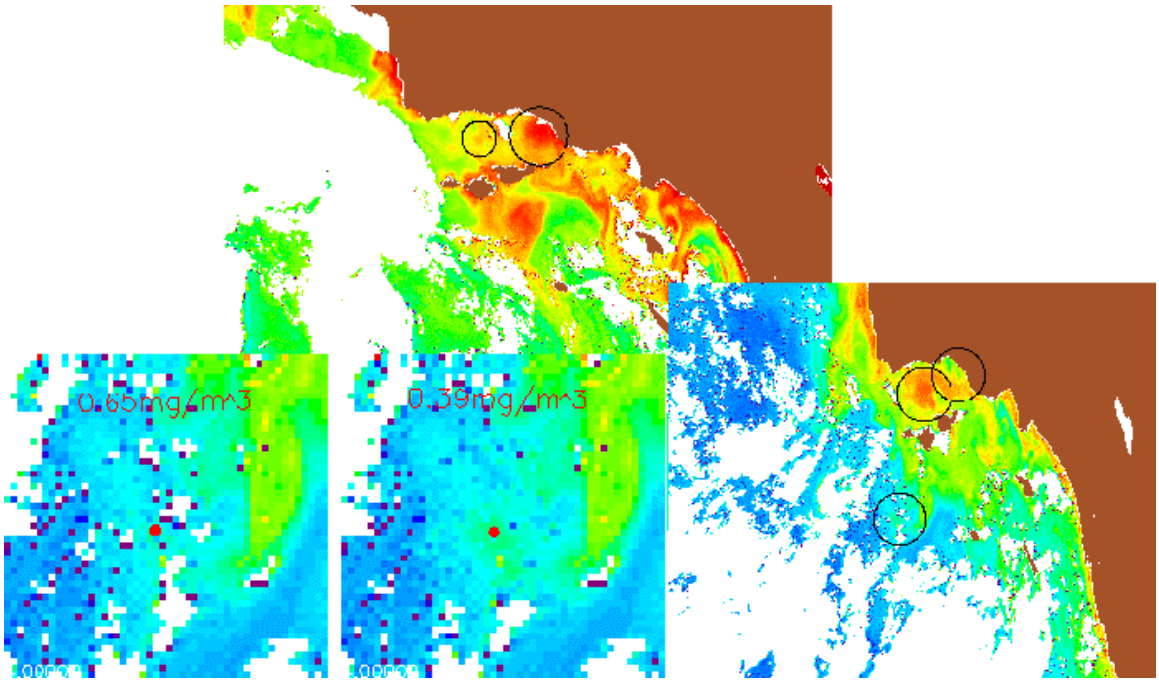



Fig. 14. Merger of CalCOFI *in situ* chlorophyll measurements with SeaWiFS L2 data using the wavelet multiresolution approach. The option of filling small cloud gaps in ocean color satellite data was applied in the bottom left-hand side scene.

*4.3 Conclusions*

A tool was implemented to investigate differences and local-field distribution changes in ocean color imagery when overlaying *in situ* data onto satellite scenes. The implementation was based on wavelet multiresolution analysis. The wavelets enabled the spatial spread of *in situ* values onto the imagery without smoothing the ocean color fields. Spatial variability and chlorophyll structures were also preserved. The merger application was designed to be ultimately dependent on local area spatial and temporal variabilities and data correlations. During the study, it was determined that concurrent *in situ* and satellite observations were scarce, even when data from recurrent oceanic surveys were applied. Therefore, it was intended that *in situ* data were presently used for validation of ocean color imagery and not for broad application in the merger efforts to complement satellite data.

Ainsworth, E. J. and I. S. F. Jones, 1999: Radiance Spectra Classification from the Ocean Color and Temperature Scanner on ADEOS. *IEEE Trans. Geoscience and Remote Sens.*, **37**, issue 3, 1645-1656.

Atkinson P. M. and A. R. L. Tatnall, 1997: Neural Networks in Remote Sensing. *Int. J. Remote Sens.*, **18**, no. 4, 699-709.

Azimi-Sadjadi M. R. and S. A. Zekavat, 2000: Cloud Classification Using Support Vector Machines. *Proceedings of the IEEE Geoscience and Remote Sensing Symposium*, **2**, 669-671.

Bailey, S. W., J. P. Werdell, and C. R. McClain, 2001: Validation of Satellite-Derived Ocean Color: Theory and Practice. *Eos Trans. American Geophysical Union*, Fall Meeting, San Francisco, USA, **82**, no. 47, F676 (OS52A-0520).

Barnes, R. A., R. E. Eplee Jr., G. M. Schmidt, F. S. Patt, and C. R. McClain, 2001: Calibration of SeaWiFS. I. Direct Techniques. *Applied Optics*, **40**, no. 36, 6682-6700.

Barnes, S. L., 1964: A Technique for Minimizing Details in Numerical Weather Map Analysis. *J. Applied Meteorology*, **3**, 396-410.

Blanc, P., T. Blu, T. Ranchin, L. Wald, and R. Aloisi, 1998: Using Iterated Rational Filter Banks Within the ARSIS Concept for Producing 10m Landsat Multispectral Images. *Int. J. Remote Sens*., **19**, no. 12, 2331-2343.

Brown, M., H. G. Lewis, and S. R. Gunn, 2000: Linear Spectral Mixture Models and Support Vector Machines for Remote Sensing. *IEEE Trans. Geoscience and Remote Sens.*, **38**, no. 5, 2346-2360.

Campbell, J. W., 1995: The Lognormal Distribution as a Model for Bio-optical Variability in the Sea. *J. Geophys. Research*, **100**, no. C7, 13,237-13,254.

Chomko R. M. and H. R. Gordon, 1998: Atmospheric Correction of Ocean Color Imagery: Use of the Junge Power-Law Aerosol Size Distribution with Variable Refractive Index to Handle Aerosol Absorption. *Applied Optics*, **37**, no. 24, 5560-5572.

Cristianini N. and J. Shawe-Taylor, 2000: *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge University Press.

Donoho, D. L., 1995: De-Noising by Soft-Thresholding. *IEEE Trans. on Information Theory*, **41**, no. 3, 613-627.

Eplee Jr., R. E., W. D. Robinson, S. W. Bailey, D. K. Clark, P. J. Werdell, M. Wang, R. A Barnes, and C. R. McClain, 2001: Calibration of SeaWiFS. II. Vicarious Techniques. *Applied Optics*, **40**, no. 36, 6701-6718.

Esaias, W. E., M. R. Abbot, I. Barton, O. B. Brown, J. W. Campbell, K. L. Carder, D. K. Clark, R. L. Evans, F. E. Hoge, H. R. Gordon, W. P. Balch, R. Letelier, and P. J. Minnett, 1998: An Overview of MODIS Capabilities for Ocean Science Observations. *IEEE Trans. Geosci. and Remote Sens*., **36**, no. 4, 1250-1265.

Franz B. A. and Y. Kim, 2001: A Comparative Study and Intercalibration Between OSMI and SeaWiFS. *Eos Trans. American Geophysical Union*, Fall Meeting, San Francisco, USA, **82**, no. 47, F661 (OS42D-09).

Fukuda S. and H. Hirosawa, 2001: Polarimetric SAR Image Classification Using Support Vector Machines. *IEICE Trans. Electron*., **E84-C**, no. 12.

Goldberg, D. E., 1989: *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley.

Gordon H. R. and M. Wang, 1994: Retrieval of Water-Leaving Radiance and Aerosol Optical Thickness over the Oceans with SeaWiFS: a Preliminary Algorithm. *Applied Optics*, **33**, 443-452.

Gordon, H. R., T. Du, and T. Zhang, 1997: Remote Sensing of Ocean and Aerosol Properties: Resolving the Issue of Aerosol Absorption. *Applied Optics*, **36**, no. 33, 8670-8684.

Gregg, W. W., W. E. Esaias, G. C. Feldman, R. Frouin, S. B. Hooker, C. R. McClain, R. H. Woodward, 1998: Coverage Opportunities for Global Ocean Color in a Multimission Era. *IEEE Trans. Geosci. and Remote Sens*., **36**, no. 5, 1620-1627.

Gregg W. W. and R. H. Woodward, 1998: Improvements in coverage Frequency of Ocean Color: combining Data from SeaWiFS and MODIS. *IEEE Trans. Geosci. and Remote Sens*., **36**, no. 4, 1350-1353.

Gregg W. W. and M. E. Conkright, 2001: Global Seasonal Climatologies of Ocean Chlorophyll: Blending In Situ and Satellite Data for the CZCS Era. *J. Geoph. Research – Oceans*, **106(C2)**, 2499-2515.

Gross, L., S. Thiria, R. Frouin, and B. G. Mitchell, 2000: Artificial Neural Networks for Modeling the Transfer Function Between Marine Reflectance and Phytoplankton Concentration. *J. Geophysical Research*, **105**, no. C2, 3483-3495.

Gualtieri, J. A., S. R. Chettri, R. F. Cromp, and L. F. Johnson, 1999: Support Vector Machine Classifiers as Applied to AVIRIS Data. *Summaries of the Eight JPL Airborne Earth Science Workshop*, USA.

Hearst, M. A., 1998: Support Vector Machines. *IEEE Intelligent Systems*, Jul./Aug., 18-28.

Julian P. R. and H. J. Thiebaux, 1975: On some Properties of Correlation Functions Used in Optimum Interpolation Schemes. *Monthly Weather Review*, **103**, 605-616.

Kilpatrick, K., E. Kearns, E. J. Kwiatkowska-Ainsworth, and R. L. Evans, 2002: Time Series of Calibrated Ocean Products from NASA's Moderate Resolution Scanning Spectrometer (MODIS). *Proceedings of the Ocean Sciences Meeting*, Feb., Honolulu, Hawaii, USA.

Kwiatkowska-Ainsworth, E. J., 2001: Merger of Ocean Color Information of Different Spatial Resolution: SeaWiFS and MOS. *Eos Trans. American Geophysical Union*, Fall Meeting, San Francisco, USA, **82**, no. 47, F675 (OS52A-0514).

Kwiatkowska E. J. and G. S. Fargion, 2002a: Merger of Ocean Color Information from Multiple Satellite Missions under the NASA SIMBIOS Project Office. *Proceedings of the Fifth International Conference on Information Fusion*, Jul., Annapolis, MD, USA, **1**, 291-298.

Kwiatkowska E. J. and G. S. Fargion, 2002b: Merger of Ocean Color Data from Multiple Satellite Missions within the SIMBIOS Project. *Proceedings of SPIE Symposium – Remote Sensing of the Atmosphere, Ocean, Environment, and Space*, Oct., Hangzhou, China, Ocean Remote Sensing and Applications, **4892**, 168-182.

Kwiatkowska, E. J., 2003: Application of Machine Learning Techniques Towards the Creation of a Consistent and Calibrated Global Chlorophyll Concentration Baseline Dataset Using Remotely Sensed Ocean Color Data. *IEEE Trans. Geosci. and Remote Sens*., accepted.

Linde, Y., A. Buzo, and R. M. Gray, 1980: An Algorithm for Vector Quantizer Design. in IEEE Trans. Commun., COM-28, **1**, 84-95.

Mallat, S. G., 1989: A Theory for Multiresolution Signal Decomposition: The Wavelet Representation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **11**, no. 7, 674-693.

Maritorena, S., D. A. Siegel, A.R. Peterson and M. Lorenzi-Kayser, 2000: Tuning of a Pseudo-analytical Ocean Color Algorithm for Studies at Global Scales. *Eos Trans. American Geophysical Union*, Jan., San Antonio, USA, OS12M-05.

Maritorena, S., D. A. Siegel, and A. Peterson, 2002: Optimization of a Semi-Analytical Ocean Color Model for Global Scale Applications. *Applied Optics*, **41**, no. 15, 2705-2714.

Masters, T., 1994: *Signal and Image Processing with Neural Networks. A C++ Sourcebook*. John Wiley & Sons, inc.

Núñez, J., X. Otazu, O. Fors, A. Prades, V. Palà, R. Arbiol, 1999: Multiresolution-Based Image Fusion with Additive Wavelet Decomposition. *IEEE Trans. Geosci. and Remote Sens.*, **37**, no.3, 1204-1211.

O'Reilly, J. E., S. Maritorena, B. G. Mitchell, D. A. Siegel, K. L. Carder, S. A. Garver, M. Kahru, and C. R. McClain, 1998: Ocean Color Chlorophyll Algorithms for SeaWiFS. *J. Geoph. Research*, **103**, no. C11, 24,937-24,953.

Pao, Y.-H., 1989: *Adaptive Pattern Recognition and Neural Networks*. Reading, MA, Addison-Wesley.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, 1992: *Numerical Recipes in C. The Art of Scientific Computing*. Cambridge University Press.

Reynolds, R. W., 1988: A Real-Time Global Sea Surface Temperature Analysis. *J. Climate*, **1**, 75-86.

Reynolds R. W. and T. M. Smith, 1994: Improved global sea surface temperature analyses using optimum interpolation. *J. Climate*, **7**, 929-948.

Rioul O. and M. Vetterli, 1991: Wavelets and Signal Processing. *IEEE SP Magazine*, 14-38.

Schölkopf, B., K. Sung, C. Burges, F. Girosi, P. Niyogi, T. Poggio, and V. Vapnik, 1996: Comparing Support Vector Machines with Gaussian Kernels to Radial Basis Function Classifiers. *Massachusetts Institute of Technology Artificial Intelligence Laboratory Memorandum*, **1599**, C.B.C.L. paper no. 142.

Schölkopf, B., 2000: Statistical Learning and Kernel Methods. *Microsoft Research Technical Report*, no. MSR-TR-2000-23, 1-27.

Starck J.-L. and F. Murtagh, 1994: Image Restoration with Noise Supression Using The Wavelet Transform. *Astron. Astrophys.*, **288**, 342-348.

Tanaka, A., H. Kobayashi, M. Kishino, T. Oishi, R. Doerffer, H. Schiller, and T. Kubota, 2000: Neural Network Algorithm for Simultaneious Retrieval of Aerosol and Water Constituents Based on Atmospshere-Ocean Coupled Radiative Transfer Model, *Proceedings of the Ocean Optics Conference*, no. XV.

Thiebaux, H. J., 1973: Maximally Stable Estimation of Meteorological Parameters at Grid Points. *J. Atmospheric Science*, **30**, 1710-1714.

Thiebaux H. J. and M.A. Pedder, 1987: *Spatial Objective Analysis with Applications in atmospheric Sciences*. Academic Press.

Wang M. and B. A. Franz, 2000: Comparing the Ocean Color Measurements between MOS and SeaWiFS: A Vicarious Intercalibration Approach for MOS. *IEEE Trans. Geosci. and Remote Sens*., **38**, no. 1, 184-197.

Wang, M., A. Isaacman, B. A. Franz, and C. R. McClain, 2002: Ocean-Color Optical Property Data Derivation from the Japanese Ocean Color and Temperature Scanner and the French Polarization and Directionality of the Earth's Reflectances: a Comparison Study. *Applied Optics*, **41**, no. 6, 974-990.

Werdell J. P. and S. W. Bailey, 2002: The SeaWiFS Bio-optical Archive and Storage System (SeaBASS): Current Architecture and Implementation. *NASA Technical Memorandum*, no. 2002-211617, G. S. Fargion and C. R. McClain, Eds.